Year-round distribution of Northeast Atlantic seabird populations: applications for population management and marine spatial planning

Per Fauchald[1,]*, Arnaud Tarroux[1], Françoise Amélineau[2], Vegard Sandøy Bråthen[3], Sébastien Descamps[2], Morten Ekker[4], Halfdan Helgi Helgason[2,21], Malin Kjellstadli Johansen[2], Benjamin Merkel[2,5], Børge Moe[3], Jens Åström[3], Tycho Anker-Nilssen[3], Oskar Bjørnstad[6], Olivier Chastel[7], Signe Christensen-Dalsgaard[3], Jóhannis Danielsen[8], Francis Daunt[9], Nina Dehnhard[3], Kjell Einar Erikstad[1], Alexey Ezhov[10], Maria Gavrilo[11,12,22], Gunnar Thor Hallgrimsson[13], Erpur Snær Hansen[14], Mike Harris[9], Morten Helberg[15], Jón Einar Jónsson[16], Yann Kolbeinsson[17], Yuri Krasnov[10], Magdalene Langset[3], Svein-Håkon Lorentsen[3], Erlend Lorentzen[2], Mark Newell[9], Bergur Olsen[8], Tone Kristin Reiertsen[1], Geir Helge Systad[18], Paul Thompson[19], Thorkell Lindberg Thórarinsson[17], Sarah Wanless[9], Katarzyna Wojczulanis-Jakubas[20], Hallvard Strøm

[1]Norwegian Institute for Nature Research, Fram Centre, 9296 Tromsø, Norway

Full author addresses are given in the Appendix

ABSTRACT: Tracking data of marine predators are increasingly used in marine spatial management. We developed a spatial data set with estimates of the monthly distribution of 6 pelagic seabird species breeding in the Northeast Atlantic. The data set was based on year-round global location sensor (GLS) tracking data of 2356 adult seabirds from 2006–2019 from a network of seabird colonies, data describing the physical environment and data on seabird population sizes. Tracking and environmental data were combined in monthly species distribution models (SDMs). Cross-validations were used to assess the transferability of models between years and breeding locations. The analyses showed that birds from colonies close to each other (<500 km apart) used the same nonbreeding habitats, while birds from distant colonies (>1000 km) used colony-specific and, in many cases, non-overlapping habitats. Based on these results, the SDM from the nearest model colony was used to predict the distribution of all seabird colonies lying within a species-specific cut-off distance (400–500 km). Uncertainties in the predictions were estimated by cluster bootstrap sampling. The resulting data set consisted of 4692 map layers, each layer predicting the densities of birds from a given species, colony and month across the North Atlantic. This data set represents the annual distribution of 23.5 million adult pelagic seabirds, or 87% of the Northeast Atlantic breeding population of the study species. We show how the data set can be used in population and spatial management applications, including the detection of population-specific non-breeding habitats and identifying populations influenced by marine protected areas.

KEY WORDS: *Fulmarus glacialis · Rissa tridactyla · Uria aalge · Uria lomvia · Alle alle · Fratercula arctica* · Marine spatial planning · SEATRACK

---

1. INTRODUCTION

Maps of the spatial distribution of endangered and vulnerable marine animals are powerful tools in eco-

system-based marine spatial planning for the creation of marine protected areas (MPAs) and in environmental impact assessments of human activities such as fisheries, shipping, oil and gas exploitation

*Corresponding author: per.fauchald@nina.no

and offshore wind farms (Lascelles et al. 2012, Harrison et al. 2018, Hays et al. 2019). Traditionally, spatial data are collected on surveys where organisms are counted and mapped in a given area. For migratory animals, the provenance, or origin, of the observed individuals is often unknown (Perrow et al. 2015, Sansom et al. 2018), and for survey data, it might be difficult to link a spatial environmental risk or management measure to the potential impacts on population dynamics. For example, without knowledge of provenance, it can be challenging to identify the breeding populations affected by an oil spill that kills seabirds at sea during the nonbreeding period (e.g. Cadiou et al. 2004). Similarly, it can be difficult to assess the population consequences of bycatch in offshore fisheries of e.g. seabirds, sea turtles and marine mammals.

The missing link between space and population can be resolved by tagging individuals at the breeding site and tracking them during the nonbreeding period. In this way, it is possible to obtain data that will help identify population-specific migratory pathways and staging and wintering areas, thus shifting the focus from an area-centred to a population-centred approach. The population-centred approach is particularly useful for animals roaming over large areas that would otherwise be difficult to cover by traditional surveys (see e.g. Block et al. 2011, Lascelles et al. 2016, Carneiro et al. 2020).

However, animals from different breeding populations might use different staging and wintering areas (e.g. Fayet et al. 2017, Merkel et al. 2021a). Accordingly, when the goal is to identify important areas for marine megafauna, or 'hotspots' at sea (e.g. Block et al. 2011, Lascelles et al. 2016, Hindell et al. 2020), it is necessary to obtain a balanced sample of all breeding populations that could potentially use the area (i.e. Aarts et al. 2008, Péron et al. 2018). In other words, a population-centred approach requires a spatially explicit design with respect to the populations covered.

Building on the trove of positional data generated through recent methodological and technical advances, species distribution models (SDMs) have become a widely used tool for predicting and mapping the distribution of flora and fauna (Guisan & Thuiller 2005, Elith & Leathwick 2009). SDMs

are used in various management applications such as conservation planning of endangered species, impact assessments of human activities and spatial planning of protected areas (Guisan & Thuiller 2005, Melo-Merino et al. 2020). Briefly, SDMs are empirical models that relate species occurrence data to environmental predictors (Guisan & Zimmermann 2000). The relationship can be estimated by various statistical methods and is expected to reflect the environmental niche utilized by the species. If the realized niche is constant across space, the relationship can be used to predict the spatial distribution of the species in areas where the environmental variables are known. Given the environmental conditions, the resulting model predictions represent a quantitative estimate of the distribution of a species and are frequently used to inform decision makers in management processes (Guisan & Thuiller 2005, Elith & Leathwick 2009).

Drawing on seabird population monitoring programs in the UK, Faroe Islands, Iceland, Russia and Norway, the SEATRACK project aims to track seabirds from a representative sample of seabird colonies in the Northeast Atlantic using global location sensor (GLS) loggers (Strøm et al. 2021, this Theme Section) (Fig. 1). In combination with data on seabird colony sizes across the study area, these tracking data provide a unique opportunity to model not only habitat
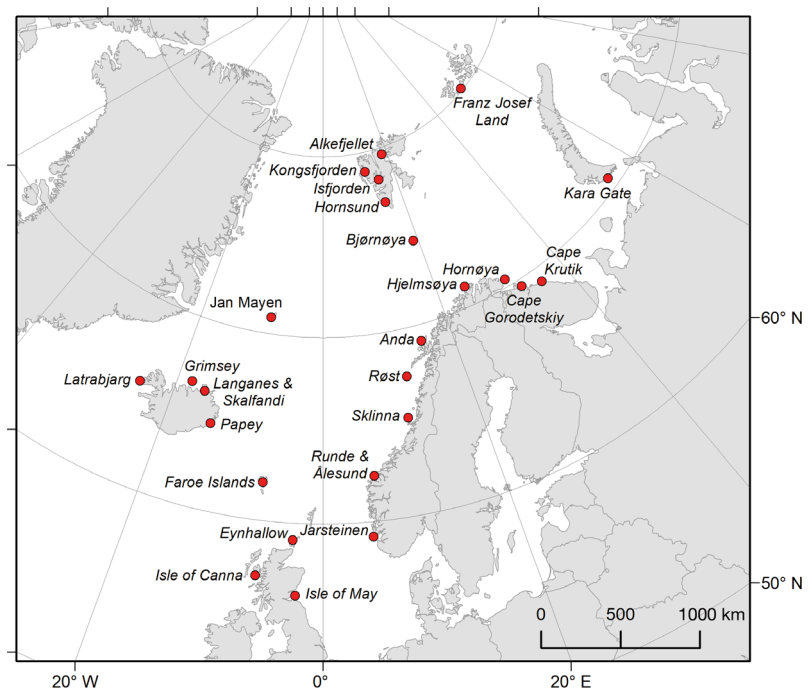

Fig. 1. Seabird colonies with seabird position data sets (GLS data) used to model the distribution of Northeast Atlantic seabirds

utilization but also extend the modelling approach to model seabird abundances for populations across the Northeast Atlantic for multiple seabird species.

In this study, we combined position data from the SEATRACK database with data on the marine environment and seabird population data to model the year-round spatial distribution of Northeast Atlantic populations of 6 pelagic seabird species: northern fulmar *Fulmarus glacialis*, black-legged kittiwake *Rissa tridactyla*, common murre *Uria aalge*, thick-billed murre *Uria lomvia*, little auk *Alle alle* and Atlantic puffin *Fratercula arctica*. These species represent the most numerous seabird species breeding in the Northeast Atlantic (Barrett et al. 2006, Frederiksen 2010). However, due to recent population declines, several of the species are listed in the European Red List of Birds (BirdLife International 2015). These are the northern fulmar (endangered), black-legged kittiwake (vulnerable) common murre (near threatened) and Atlantic puffin (endangered). While all 6 species are long-lived pelagic seabirds, they dif-

fer in trophic level and foraging mode (Del Hoyo et al. 1992), and it is therefore expected that they will utilize different habitats and show different responses to the marine environment and potential stressors.

The workflow of our analyses is shown in Fig. 2. Position data from the SEATRACK database were combined with environmental data to model the monthly habitat use of the species from each SEATRACK colony. Environmental variables were derived from data on bathymetry and remote sensing data of sea surface temperature (SST), sea surface height and primary production. We used cross-validation to assess how the models performed when predicting the spatial distribution of birds in other years and birds from other colonies, and we used cluster bootstrap sampling to estimate the CIs of the predictions. To link the habitat models to seabird populations, we compiled a colony database for the Northeast Atlantic and conducted spatial analyses to justify distance rules for assigning the populations in the colony database to the nearest modelled seabird
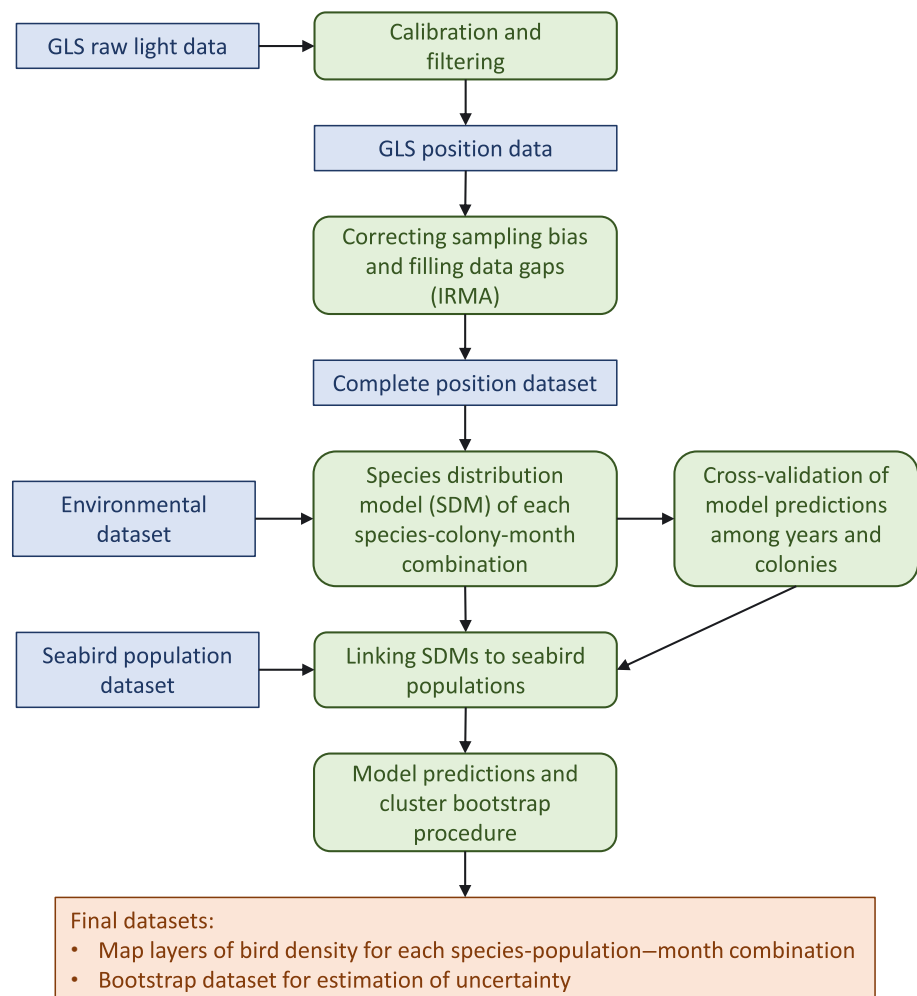
Fig. 2. Workflow, including input data sets (blue boxes) and analyses (green boxes) to obtain a spatial data set with estimates of the monthly distribution of 6 pelagic seabird species breeding in the Northeast Atlantic. IRMA: informed random movement algorithm

GLS raw light data → Calibration and filtering

GLS position data

Correcting sampling bias and filling data gaps (IRMA)

Complete position dataset

Environmental dataset → Species distribution model (SDM) of each species-colony-month combination → Cross-validation of model predictions among years and colonies

Seabird population dataset → Linking SDMs to seabird populations

Model predictions and cluster bootstrap procedure

Final datasets:
• Map layers of bird density for each species-population—month combination
• Bootstrap dataset for estimation of uncertainty

colony. Based on the distance rules, we predicted the habitat use of each colony covered by the SEATRACK design. Finally, we weighted the estimates with population size, yielding monthly maps of the abundance of Northeast Atlantic breeding seabirds. We show how the maps of the predicted distributions can be used in (1) population management applications, by identifying the year-round distribution of vulnerable or threatened populations; and (2) spatial management applications, by identifying the populations potentially affected by area-specific human stressors or management efforts.

2. MATERIALS AND METHODS

The study area covered the North Atlantic Ocean from 78° W to 80° E and from 35–85° N. Position data sets were given as time and geographical coordinates. Environmental variables were aggregated on a 0.25 × 0.25° geographic grid covering the study area, and model predictions were conducted on the same grid. All distances used in the analyses were calculated as great-circle distances in km, and when relevant, the distances were calculated around land masses. Maps presented are shown in a stereographic projection with the origin at 90° N and 0° E. All analyses were conducted in R v.3.6.3 (R Development Core Team 2020).

2.1. Data sets

2.1.1. GLS data

GLS data from 6 species of pelagic seabirds (northern fulmar, black-legged kittiwake, common murre, thick-billed murre, little auk and Atlantic puffin) were obtained from 25 seabird colonies (Fig. 1) from 2006–2019. Note that data before 2014 were collected as part of projects other than the SEATRACK project. The number of tracks, birds, years with data and positions are given in Table S1 in the Supplement at www.int-res.com/articles/suppl/m676p255_supp.pdf for each species, colony and month. See Table 1 for overall sample sizes for each species. In total, the data set comprised 2356 tracked birds, resulting in 4827 bird-tracking years. Adult breeding birds were equipped during the breeding season and recaptured to retrieve the logger in following breeding seasons. The raw light data yielded up to 2 locations ind.$^{-1}$ d^{-1} (12:00–00:00 h). The positions obtained from light measurements are prone to a num-

ber of errors. First, latitude cannot be estimated adequately within a few weeks of the equinox in autumn and spring, when the day length is virtually the same everywhere on Earth. Second, neither latitude nor longitude can be determined in Arctic areas during polar day and polar night, when the diurnal variation in light level is insufficient to be detected by the light loggers. Finally, shading due to behaviour or weather can prevent the loggers from recording twilight events properly. A threshold method was used to calculate positions from the light data (Lisovski et al. 2020); next, raw positions were filtered with speed, distribution, angle, distance and loess filters. The protocol for obtaining positions from the raw light measurement, including calibration of different GLS models and filtering procedures, is described in Bråthen et al. (2021). In addition to the light sensor, the GLS loggers include a salinity sensor ('wet/dry') that records periods when the logger is submerged in saltwater. These data were used to determine when the birds departed and arrived at the breeding colony and whether they were mostly sitting on their nest (dry) or at sea (wet) during the breeding season (see below). The precision of the GLS tracks is relatively low and likely to change throughout the tracking period (Lisovski et al. 2020). Thus, it is only possible to identify habitat use on a relatively coarse spatial scale (i.e. 100s of km) (Phillips et al. 2004).

2.1.2. Environmental data

Environmental data were used as predictors in the SDMs. Ideally, the predictors should encompass features with a direct link to the species' habitat selection. However, the available predictors were limited to data sets covering the entire study area. For dynamic environmental variables, these were restricted to parameters measured by remote sensing which, in many cases, will serve as proxies for more direct drivers of habitat selection such as food availability. Specification of the environmental variables and data sources are given in Table S2. SST (*SST*) is an important parameter often used to differentiate water masses and to model the distribution of seabirds (Tremblay et al. 2009). We used monthly values of *SST* as predictors in the SDMs. Seabirds are often attracted to frontal areas where marine production is concentrated (Fauchald 2009). We calculated the spatial variation in *SST* (*Front*) and used this measure as a proxy for frontal areas between water masses. Meso-scale eddies are oceanographic features that

can influence the distribution of seabirds (Hyrenbach et al. 2006). We used the average monthly sea surface level, i.e. sea surface level above geoid (*Adt*), to reflect meso-scale eddies in the study area. Areas with high primary production could indirectly influence seabird distribution through enhanced food availability. We used the integrated annual primary production (*Prim*) to reflect areas of high primary production. S tatic environmental variables included in the SDMs were bottom depth (*Depth*), indicating features such a s seamounts and shelves; spatial variance in *Depth* (*Slope*), indicating important bathymetric features such as the continental edge affecting currents and delineating water masses; and distance to coast (*CoastD*), reflecting terrestrial input and influence. Finally, seabirds are visual predators, and we hypothesized that short daylength could be a limiting factor in the northernmost areas during the darkest winter months. Daylength (*Daylen*) was therefore included as a predictor from November through January. We assumed that the seabirds did not use heavily sea-ice covered areas, and we compiled monthly data on sea ice to mask areas with sea ice concentration >80%. All predictors were adjusted to fit the spatial raster covering the study area and with a pixel resolution of 0.25 × 0.25°. Data with a lower spatial resolution were either disaggregated or re-sampled with a bilinear local interpolation using the 'raster' package v.3.4-5 (Hijmans 2020).

2.1.3. Seabird population data

Population count data were compiled from different sources to generate a representative data set of the breeding populations of the 6 seabird species in the Northeast Atlantic. Data sources are given in Table S3. Note that in the present definition of Northeast Atlantic populations, we excluded the relatively small seabird populations from the Baltic Sea, Kattegat, Germany, France and the southern portions of the UK and Ireland. The spatial resolution of the defined 'colonies' differed among data sources, and for several regions, the data set comprised a large number of small proximate breeding sites. To obtain a homogeneous and manageable data set, the original colony data set was aggregated using a spatial hierarchical cluster analysis using the 'hclust' procedure from 'stats' package v.3.6.3. A cut-off distance equal to 50 km was used in the clustering procedure. The clustering reduced the number of colonies (site–species combinations) from 2315 to 689.

2.2. Correcting for sampling biases

Data gathered from tracking devices are presence-only data. This means that an observed location represents a 'true' presence of an individual, but the data set gives no information of the contrast, i.e. observations of 'true' absences of individuals. As a consequence, any sampling bias in the tracking data is directly translated into a bias in the estimated spatial distribution of the population. For example, gaps in the positions obtained from the GLS tags due to stable light conditions during polar night or polar day will bias the presence data set to the south, out of Arctic areas with continuous darkness or daylight. In other words, non-random gaps in the data set will contribute to sampling biases that will affect the interpretation of the spatial distribution. Such gaps include the lack of registration of positions during equinoxes and polar night/day or removal of 'false' positions over land and fast sea-ice. To alleviate these biases, we used an informed random movement algorithm (IRMA) to model new locations to fill in these non-random gaps (Fauchald et al. 2019). IRMA builds on a method originally proposed by Technitis et al. (2015) to generate plausible locations using additional information available: (1) conductivity data, to determine attendance at the colony; (2) land masks, to constrain positions over the ocean; (3) longitudes (that can be estimated during the equinoxes) and (4) light levels, indicating if the birds were north or south of the Arctic Circle (i.e. continuous day/night recorded by the loggers). In addition, because we focused on pelagic species, we assumed that all GLS positions occurring over land areas were wrong, and the locations were thus removed before running IRMA through the entire data set.

2.3. SDMs

Several modelling techniques are available for fitting SDMs from presence-only data (see e.g. Guisan & Thuiller 2005, Elith & Leathwick 2009). These include regression techniques such as generalized linear models (GLMs) and generalized additive models (GAMs) and machine learning techniques such as boosted regression trees and maximum entropy. Ensemble models that combine the output from several modelling techniques have also been advocated (Scales et al. 2016). In the present study, we used the GAM technique (Hastie & Tibshirani 1990, Wood 2017). GAM is a well-proven and computational efficient regression method that relates a linear response

variable to smooth functions of predictor variables. GAM offers a straightforward interpretation of the fitted model and regularization of the predictor function to avoid overfitting, and comparisons suggest its performance is similar to machine learning techniques (see Wisz et al. 2008, Elith et al. 2010, Scales et al. 2016, Norberg et al. 2019).

In SDMs of presence-only data, it is necessary to introduce background points to contrast the recorded presences in the analyses (Elith & Leathwick 2009, Barbet-Massin et al. 2012). It is essential that the background data constitutes a representative set of environmental conditions available to the species (Phillips et al. 2009). In the present analyses, background data points were selected randomly from an area defined by the minimum convex polygon of all positions of the species (including IRMA positions) using the 'mcp' function in the 'adehabitatHR' package in R (Calenge 2006). To the polygon, we added a buffer with a width equal to 10% of the radius of a circle defined by the area of the minimum convex polygon. We assumed that the seabirds did not occur over land and fast sea-ice, and we accordingly removed areas covered by land and areas with more than 80% sea-ice concentration. For each model, we drew 20 000 background points from the defined potential habitat and extracted the relevant environmental data at each location. SDMs of presence-background data for each species, colony and month were fitted using the 'gam' function in the 'mgcv' package v.1.8-33 (Wood 2017). The probability of presence/background was modelled using a logit-link function with a binomial distribution.

2.3.1. Dynamic and static predictors

Our purpose was to model the monthly distribution of each species from each colony. Accordingly, data from all available years were combined in separate SDMs for each species, colony and month. Because year-to-year changes in the marine environment were expected to affect the spatial distribution of birds, environmental data were assigned to the presence data by matching position, year and month. Of the environmental variables included in the models, *SST*, *Front* and *Adt* varied by month and year, *Prim* varied by year, *Daylen* varied by month only, and *Depth*, *Slope* and *CoastD* were static variables (Table S2). Year-to-year changes in the environment were accordingly accounted for by SST, frontal areas, sea surface height and primary production. To account for the environmental dynamics in the selection of background points, the environmental variables in the potential habitats were extracted for each month and year, giving monthly data sets of potential habitats. In order to combine background points from different years in the models, the number of background points drawn from a potential habitat was proportional to the number of presences in that year, keeping the total number of background points equal to 20 000. According to this procedure, the models estimate the average response to the environment corresponding to the sample size in each year. Note that the models were (1) evaluated according to how they predicted the distribution of birds in out-of-sample years (Section 2.4), and (2) the habitat predictions given in the final data set are based on the environment in 2018 (Section 2.7), a year that was well represented in the sample.

2.3.2. Constraining predictors

The habitat used by a population of seabirds is the result of a combination of the birds' selection of specific environmental features and the birds' spatial constraints. One obvious spatial constraint is the distance to the breeding colony, which will have a seasonally shifting impact on habitat use depending on the birds' behavioural bonds to the colony. Distance to colony (*ColD*) was therefore included as a predictor in the models. Other spatial constraints could be inherited or learned spatial behaviour, including fixed migration routes or wintering areas (Merkel et al. 2021a, Léandri-Breton et al. 2021, this Theme Section). Studies show that the seabird species in our study area have colony-specific migration routes and wintering areas (Frederiksen et al. 2012, 2016, Fayet et al. 2017, Merkel et al. 2021b, this Theme Section, Amélineau et al. 2021, this Theme Section), indicating that such constraints are important in our study system (Léandri-Breton et al. 2021, Merkel et al. 2021a). The east–west direction is particularly important because the seasonal migration of seabirds is generally dominated by an east–west component, and as a consequence, several initial models predicted false positive presences in the far east or west of the available habitat. Therefore, to better model colony-specific habitat use in the east–west direction, we included longitude (*Easting*) as a predictor in the models.

2.3.3. Spatial dependencies

One important problem related to SDMs of animal track data is the spatial dependencies, or serial auto-

correlation, among positions of the same individual (Aarts et al. 2008). This problem arises because positions that are near in time on a continuous track will also be geographically close to each other and have similar environmental attributes. Such spatial dependencies will increase the tendency of model over-fitting and the detection of spurious relationships with environmental variables. To alleviate this problem, Aarts et al. (2008) suggested including individual tracks as a random component in a mixed-model setting. This was computationally unfeasible for our data set, and to reduce the impact from over-fitting, we limited the complexity of the models by restricting the number of possible non-linear relationships. We assumed linear relationships for variables where we expected a dominant monotonic, either increasing or decreasing relationship with the presence of birds. These variables included *Front*, *Adt*, *Prim*, *Depth*, *Slope, Coastd* and *Daylen*. Daylength was only included in the models during the 3 darkest months (November–January). We assumed that the relationship between presences and *ColD*, *Easting* and *SST* were likely non-linear and modelled these variables with tensor product smooth functions with cubic regression splines as basis. The optimal degree of smoothing was determined by generalized cross-validation (GCV). In cases when the models did not converge, the complexity of the models was constrained by specifying a maximum number of knots in the smooth function (k = 5 or 4).

2.3.4. Multicollinearity

Several of the explanatory variables describe, at least partly, similar physical or biological features. For example, SST, daylength and sea surface height have a natural south to north gradient, resulting in large-scale positive spatial correlations among these variables (Pearson's r ranging between 0.24 and 0.72). However, removing any of these variables from the models reduced the models' ability to predict presences, and therefore we decided to include all variables in the models. Because multicollinearity inflates the variance of the parameter estimates, it is a serious problem for model inference, i.e. focusing on the model's parameter estimates. However, multicollinearity does not affect the accuracy of model predictions within the range of the data. In other words, if the main purpose of the model is to predict new cases within the range of the sampled data (i.e. to interpolate), the model will do this reliably as long as the collinearity between variables remains

constant (Dormann et al. 2013). In the present study, all predictions are done within the temporal and spatial range of the data, and collinearity among predictors will therefore have little effect on our results. Nevertheless, we calculated the variance inflation factor (VIF) of the predictors of each SDM (Fig. S1). The *ColD* variable coincided with longitude for the northeastern colonies (the Russian colonies), and the VIF values for *ColD* and *Easting* were accordingly high for models of these colonies (VIF > 10). Clearly, this collinearity would affect the interpretation of how distance to colony impacts the habitat use of birds from the northeastern colonies; however, it will not affect the predicted distribution of birds. Potentially more problematic is the high VIF values of *SST* (VIF values between 5 and 10). This is because *SST* varies between years and would be an important parameter to address when projecting future seabird distribution under climate change. The high VIF values were mainly due to collinearity with *Daylen* (Pearson's r = 0.63). Because daylength is fixed among years and could limit the northerly distribution of seabirds during winter, predictions of changed winter distribution based on projected increases in sea temperature would accordingly be problematic. The present models can therefore be used to assess the habitat use of seabirds within the sampling period, but collinearity among important static and dynamic drivers would preclude the models' ability to project the distribution of birds into a warmer future.

2.4. Model performance

The performance of habitat models is usually assessed by dividing the data set into a training data set, used to fit the model, and an independent test data set, used to evaluate the predictions from the model (i.e. cross-validation). We first evaluated the within-colony performance by withholding data from one year at a time as a test data set and used the data from the remaining years to fit the model (i.e. between-year transferability; Péron et al. 2018). The fitted model was used to predict the distribution in the test year, and model performance was evaluated by comparing model predictions and the test data set. All combinations of test and training years were pooled in the performance statistics.

Secondly, we evaluated the between-colony performance by using the model from one colony to predict the distribution of occurrences of birds from other colonies. The between-colony performance is

instrumental to evaluate the representativity of the models i.e. whether a model from one colony can be used to predict the distribution of birds from other colonies (Péron et al. 2018). The fitted model from one training colony was used to predict the distribution of birds from a test colony, and the occurrence data from the test colony was used to evaluate the performance of the model predictions. Performance statistics were calculated for all combinations of training and test colonies. We restricted the analyses to the period from September to February (winter and non-breeding in the Northern Hemisphere), a period where the birds are expected to be less spatially constrained by the colony.

We used 2 different statistics to evaluate model performance: area under the receiver operating characteristic curve (AUC-ROC) and the Boyce index. Traditionally, the performance of SDMs have been measured by the AUC-ROC (see Lobo et al. 2008). AUC-ROC measures the ability of a model to distinguish between classes. In the case of SDMs on presence-only data, it measures how well the model is able to discriminate between presence and background points. For AUC-ROC values close to 1, the classification is close to perfect; for values equal to 0.5, the model classification is not better than a random classification; and for values below 0.5, the model performs an inverse classification — it predicts presences in areas dominated by background points. The use of AUC-ROC as a measure of model performance has been criticized and is particularly questionable in models of presence-only data (Lobo et al. 2008). The major reason is that the background points do not represent 'true absences'. Accordingly, the AUC-ROC measure is sensitive to the definition of the area from where the background points were drawn. By expanding the area outside the core area of the species, the AUC-ROC value will increase. The AUC-ROC value should accordingly be assessed relative to the definition of the potential habitat. In the present study, this area was defined by the extent of all observations of the species in the data set. Accordingly, all models of the same species used the same area when drawing background points. Moreover, the same procedure was used when defining the potential habitat of all 6 species. Therefore, we argue that the AUC-ROC values obtained in this study can be compared within species and also, with some limitations, across species. It is, however, important to note that the AUC scores should be interpreted relative to the extent of the area occupied by the Northeast Atlantic populations of the species. AUC-ROC values were calculated using the 'auc' procedure in the 'spatstat' package v.1.64-1 (Baddeley et al. 2015).

Boyce et al. (2002) suggested a performance indicator based on the relationship between the number of occurrences and the corresponding binned values of suitability (Boyce index) (Hirzel et al. 2006). The relationship is measured by the Spearman rank correlation coefficient. A Boyce index value close to 1 suggests a close relationship between the number of occurrences and the predicted suitability; a value equal to 0 suggests no relationship; while a negative value suggests an inverse relationship between the predicted suitability and number of occurrences. Contrary to the AUC-ROC, the Boyce index is independent of the area of the defined available habitat and has therefore been suggested to be an appropriate measure for the performance of presence-only SDMs (Hirzel et al. 2006). We calculated the Boyce index using the moving window method suggested by (Hirzel et al. 2006) by applying the 'boyce' function in the 'ecospat' package v.3.1 (Broennimann et al. 2020).

To investigate how model performance varied with sample size (Delord et al. 2014, Lascelles et al. 2016), we fitted linear models using the AUC-ROC and Boyce index as response variables, and species, number of tracks and interaction thereof as explanatory variables. Number of tracks used to fit the models varied from 18 to 194. To attain normality of residuals, the Boyce index was logit($[x + 1] / 2$)-transformed and the AUC-ROC was logit(x)-transformed in the analyses.

2.5. Estimation of uncertainty

The spatial dependencies among positions collected on the track of an individual need to be considered when estimating the uncertainty around SDMs predictions. To meet this goal, we used a cluster bootstrap procedure to estimate CIs for the model predictions. Nonparametric bootstrapping (Efron & Tibshirani 1993, Davison & Hinkley 1997) is a robust technique that builds sampling distributions for the statistics by resampling the observations from a data set. To deal with within-individual dependencies among observations, cluster bootstrapping resamples at the level of clusters or individuals rather than at the level of single measurements (Davison & Hinkley 1997). This implies that when a subject (e.g. a track) is drawn into a bootstrap sample, all observations from the subject are part of the bootstrap sample.

For each SDM, we built 400 bootstrap samples of occurrences by resampling at the level of individual tracks. For each bootstrap sample, we fitted the corresponding GAM and predicted the suitability values in the habitat. From the 400 bootstrap habitat predictions, we estimated 95 % CIs for the predictions using the 2.5–97.5th percentile intervals for the predicted logit values.

The cluster bootstrap procedure accounts for the spatial dependencies of positions within tracks. However, the sample included birds tracked over more than 1 yr, giving more than one track for these individuals. Overall, the data set included 4827 tracks of 2356 individuals (Table 1), giving on average 2 tracking years per individual bird. If individuals show site fidelity among years, there will be a spatial dependency among tracks of the same individual. The present bootstrap procedure did not account for variation at the individual level, and individual site fidelity has been shown to be important in the study system (Léandri-Breton et al. 2021, Merkel et al. 2021a). This will reduce the effective sample size of the present study and eventually increase the CI. Thus, while the cluster bootstrap procedure gives a robust estimate of CIs based on tracks, it should be noted that individual site fidelity could increase the CI in cases when the number of tracked individuals is small (i.e. <20 individuals; cf. sample sizes in Table S1).

2.6. Assigning breeding population data to SDMs

To generate population maps from the habitat models, it was necessary to link the seabird population data set to the SDMs of the model colonies. To this end, we used the SDM from the nearest model colony. Distance between colonies was measured as the shortest ocean distance (i.e. around land masses) using the function 'gridDistance' in the 'raster' package v.3.4-5 (Hijmans 2020). A cut-off distance was used to exclude colonies not adequately covered by the spatial design of model colonies. The cut-off distance was justified by analyses of model performance among colonies (see Section 2.4). The accuracy of models in predicting the occurrences of birds from other colonies decreased with increasing distance (for Boyce index, see Fig. 3; For AUC-ROC, see Fig. S2). In general, the relationships suggested that the validity of models was high for nearby colonies and decreased and became highly variable for colonies separated by more than ca. 500 km. Clearly, deciding a cut-off distance involves a trade-off

Table 1. Samples of seabirds tracked with GLS. Number of colonies, number of tracked individuals and number of annual tracks in total and for each species

	Colonies	Individuals	Annual tracks
Northern fulmar	7	312	606
Black-legged kittiwake	16	796	1642
Common murre	9	357	844
Thick-billed murre	8	329	754
Little auk	4	194	290
Atlantic puffin	10	368	691
Total		2356	4827

between the proportion of populations covered by the design and the validity of the models (see Fig. 3). Based on these considerations, we adopted cut-off distances between 400 and 500 km, optimizing the trade-off for each species. The percentage of the Northeast Atlantic populations covered by the SEATRACK design was, according to this design, 90.6 % for northern fulmar, 90.4 % for black-legged kittiwake, 86.4 % for common murre, 83.6 % for thick-billed murre, 79.1 % for little auk and 85.5 % for Atlantic puffin.

2.7. Predictions

In the last step, the SDMs were used to predict the monthly distribution of the 6 pelagic seabird species covered by the SEATRACK design, using environmental data from 2018. While the environmental predictors were the same for all colonies, one important predictor was colony-specific, namely distance to colony. Thus, predictions had to be calculated separately for each colony. Predictions were calculated for the assigned SDM as well as for each of the corresponding 400 bootstrap SDMs giving a bootstrap sample of predictions.

Predictions were done on a 0.25° × 0.25° raster covering the habitat defined for each species (see above). SDM predictions give the relative likelihood of occurrence, and to scale these values to population size we used the following equation:

$$\hat{n}_i = N \frac{\hat{p}_i a_i}{\sum_i \hat{p}_i a_i}$$

where $\hat{n}_i$ is the predicted number of birds from a colony in raster cell i, N is the number of breeding birds in the colony (i.e. colony size), $\hat{p}_i$ is the relative likelihood of occurrence predicted by the model and a_i is the area of raster cell i.

2.8. Applications

The resulting data set comprised 4692 rasters of predicted monthly distribution of 6 pelagic seabird species from 391 colonies in the Northeast Atlantic. For each raster, a bootstrap sample was available for constructing CIs with respect to the SDM used. The data set has 2 immediate utilisations reflecting distinct management purposes that we detail below along with concrete examples.

2.8.1. Population management applications

The data set can be used to identify the seasonal habitats of specific populations. For this application, the goal would typically be to map important habitats and identify the potential threats and stressors present. Clearly, this application will be particularly important in the protection of threatened species or the management of populations of special concern. Conservation applications often involve the delineation and management of 'populations' within specific administrative or biogeographic boundaries. Within the spatial extent of the SEATRACK design, breeding populations can be defined and aggregated at any level, from the individual colony to marine ecosystems or administrative boundaries, depending on the focus of a given application. Additionally, the bootstrap samples can be used to calculate CIs for the predicted seabird distribution.

In the first example of this application, we used the data set to map the overall distribution of the combined populations of 6 Northeast Atlantic pelagic seabird species. The goal was to identify marine areas of special importance for these species. The predicted distributions of birds from each breeding colony covered by the SEATRACK design were summed and averaged in 3 mo bins, giving the combined quarterly distribution of the 6 species.

In the second example, we demonstrate how the data can be used to map the year-round distribution of a population of special concern. Due to population declines, the black-legged kittiwake is currently listed as Vulnerable on the IUCN Red List (BirdLife International 2020), and the population breeding in Norway is listed as endangered on the Norwegian Red List (Henriksen & Hillmo 2015). To estimate the spatial distribution of the Norwegian part of the population, we aggregated the estimates from all colonies on the Norwegian mainland. Bootstrap estimates were aggregated similarly and, based on the percentiles from the resulting boot-strap sample, we calculated 95% CIs of the predicted spatial distributions.

2.8.2. Spatial management applications

The data set can also be used to address threats to seabirds in marine spatial planning (e.g. Lascelles et al. 2012, 2016). Specifically, the data set can be used to identify how human activities or management actions in a given area could impact the seabird populations. The data set can provide colony-specific estimates of the number of birds spending time in the marine area in question as well as when they enter and when they leave that area, and the bootstrap estimates can be used to construct bootstrap samples for calculating corresponding CIs. It is important to note that such assessments are limited to the populations covered by the study design. It is accordingly a possibility that some important populations or species are missing from the analyses. Reducing the risk of omitting important species or populations would involve increasing the spatial extent and number of species covered by the study design, and it is recommended that the representativity of the study design is validated by independent sources of information such as expert assessments and data from surveys at sea.

Recently, a 641 612 km^2 area in the mid-Atlantic was proposed as a MPA in the Area Beyond National Jurisdiction (ABNJ) of region V as defined in the Oslo-Paris (OSPAR) convention (OSPAR 2021). The proposed area, named the 'North Atlantic Current and Evlanov Seamount' (NACE), has been identified as an important area for a number of migrating pelagic seabirds, including the 6 species covered by the present study. The importance of the area for seabirds has been documented by a designated study where a large seabird tracking data set covering 21 species was compiled and analysed (Davies et al. 2021). Here, we used the proposed delineation to identify the populations in the SEATRACK data set that use the NACE area. We calculated the seasonal habitat use and constructed CIs based on the boot-strap samples.

3. RESULTS

3.1. Model performance

In total, 648 SDMs were fitted, each representing a unique combination of species, colony and month. Model statistics (adjusted R^2 and proportion of

deviance explained) and 2 measures of model performance (Boyce index and AUC-ROC) are given for each model in Table S1. A summary of the model performance indicators for each species is given in Table 2. Both indicators increased with increasing sample size (i.e. the number of tracks in the sample: Boyce index, p = 0.001; AUC-ROC, p = 0.002). However, the increase was modest, and model performances were generally high across the range of sample sizes (Fig. S2). Therefore, we decided to keep all models in the subsequent analyses.

3.2. Model transferability

The ability of models to predict the distribution of birds from other colonies (i.e. transferability among colonies) decreased with increasing inter-colony distances (Figs. 3 & S3). Performance indicators were generally high for distances less than ca. 500 km and became highly variable for longer distances. For distances longer than ca. 1000 km, the indicators varied between values close to 1, suggesting that the model colony predicted the occurrence of birds from the test colony well, to values below 0 (below 0.5 for AUC-ROC), indicating that the model colony predicted the inverse of the distribution of birds from the test colony (Fig. 3). Thus, while colonies close to each other tended to use the same habitats, the analysis suggested that for distant colonies, the nonbreeding habitats were colony-specific and in many cases, non-overlapping. Accordingly, the SDMs could not be extrapolated to distant populations; in our cases, the limit for meaningful predictions was 400−500 km (see Section 2.6).

3.3. Model predictions

Based on the distance rules from model colonies, we predicted the spatial distribution of 391 seabird

colonies using the models from the nearest model colony. Predictions were done for each month, yielding a data set of 4692 maps. For each prediction, a bootstrap sample of 400 maps was generated. Fig. 4 shows one example of predictions and the associated 95% CI range for Atlantic puffins from Vestmannaeyjar, Iceland, in December. Vestmannaeyjar holds the world's largest Atlantic puffin colony (830 000 breeding pairs; Hansen et al. 2011). The nearest SEATRACK colony is Papey, with an ocean distance of 395 km, and the model from Papey was accordingly used to predict the distribution of birds from Vestmannaeyjar. In December, the population was, according to the model, widely distributed over a large ocean area in the Northwest Atlantic (Fig. 4A) with a maximum density of 0.6 birds km^{-2}, and 90% of the population was found within an area of 3.9 million km^2. The highest uncertainty in the predicted distribution was found along the outer periphery of the core area (Fig. 4B). Most notably, high uncertainty was present south of the core area, in the Gulf of St. Lawrence, the Grand Banks, along the Labrador coast and across the Labrador Sea to the coast of southwestern Greenland.

3.4. Population management applications

3.4.1. Seasonal distribution of Northeast Atlantic pelagic seabirds

To address important seasonal ocean areas for the 6 pelagic Northeast Atlantic seabird species, we summed the predicted densities for all populations (Fig. 5). Maps of the distribution of each of the 6 species are shown in Figs. S4−S9. The figures also indicate which colonies were encompassed by the estimates. Note in particular that colonies in southern Ireland and southern UK were not represented in the study design. During summer (Fig. 5A), roughly reflecting the breeding period for the species, the highest densities of birds were estimated close to the large colonies in Iceland, Northern Norway, Svalbard, Scotland and the Faroe Islands. The highest estimated density was 12.3 birds km^{-2}, and 90% of the birds were found within an area of 6.4 million km^2. In autumn (Fig. 5B), the birds were more widely distributed. The highest densities were found in the Barents Sea, with a peak density of

Table 2. Summary statistics of model performance indicators grouped by species. N: number of models; AUC-ROC: area under the receiver operating characteristic curve

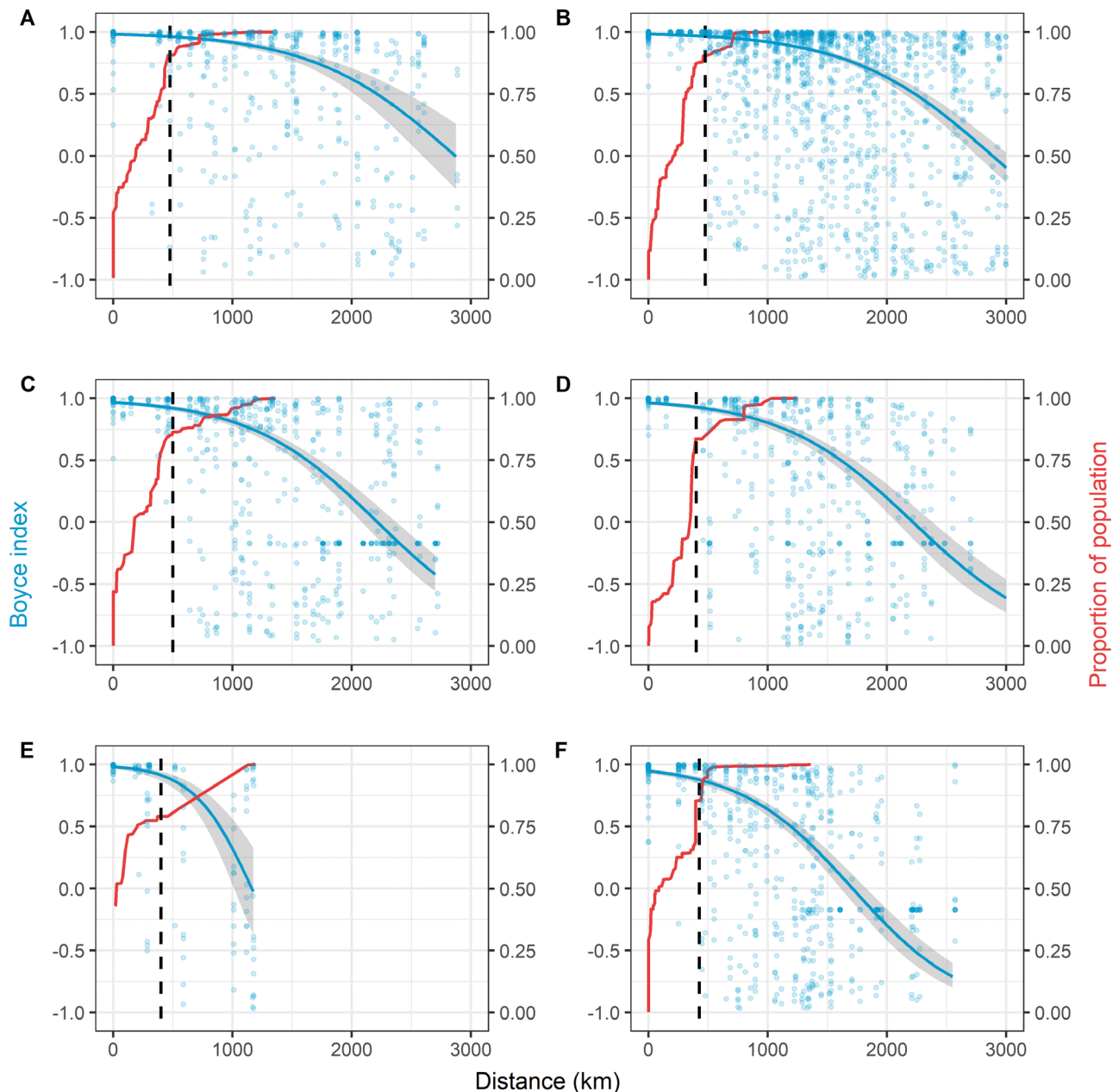|	N	Boyce index			AUC-ROC		
		Median	Min	Max	Median	Min	Max
Northern fulmar	84	0.974	0.249	1.000	0.960	0.758	0.991
Black-legged kittiwake	192	0.979	0.560	1.000	0.954	0.776	0.996
Common murre	108	0.977	0.513	1.000	0.987	0.887	0.999
Thick-billed murre	96	0.981	0.682	1.000	0.975	0.908	0.995
Little auk	48	0.925	0.375	1.000	0.921	0.823	0.977
Atlantic puffin	120	0.959	0.374	1.000	0.970	0.852	0.997

Fig. 3. Among-colony model performance (Boyce index; blue) and the cumulative proportion of the Northeast Atlantic seabird population covered by models (red) as a function of distance to model colonies for (A) northern fulmar, (b) black-legged kittiwake, (c) common murre, (d) thick-billed murre, (e) little auk and (F) Atlantic puffin. Blue points: Boyce index calculated for pairs of model and test colonies (*y*-axis) and distance between model colony and test colony (*x*-axis). Blue lines with grey shading: linear regression of Boyce index as a function of distance. The Boyce index was logit([x + 1] / 2)-transformed to attain normality; blue lines: back-transformed predictions ±95% CI. Red lines: cumulative proportion of the Northeast Atlantic populations covered by the model colonies for increasing distance to the model colonies. Hatched lines: distance limits set for predictions

9.7 birds km^{-2}, and 90% of the birds were found within an area of 8.8 million km^2. During winter (Fig. 5C), the populations spread out even more, as 90% of the birds were found within an area of 9.9 million km^2. During the darkest winter period, birds migrate out of the Barents Sea, and the highest densities were found in the Denmark Strait, between Iceland and Greenland, with a peak density of 6.9 birds km^{-2}. In spring (Fig. 5D), the distribution of birds became more concentrated again, with 90% of the birds occurring within an area of 8.7 million km^2. The highest densities were found
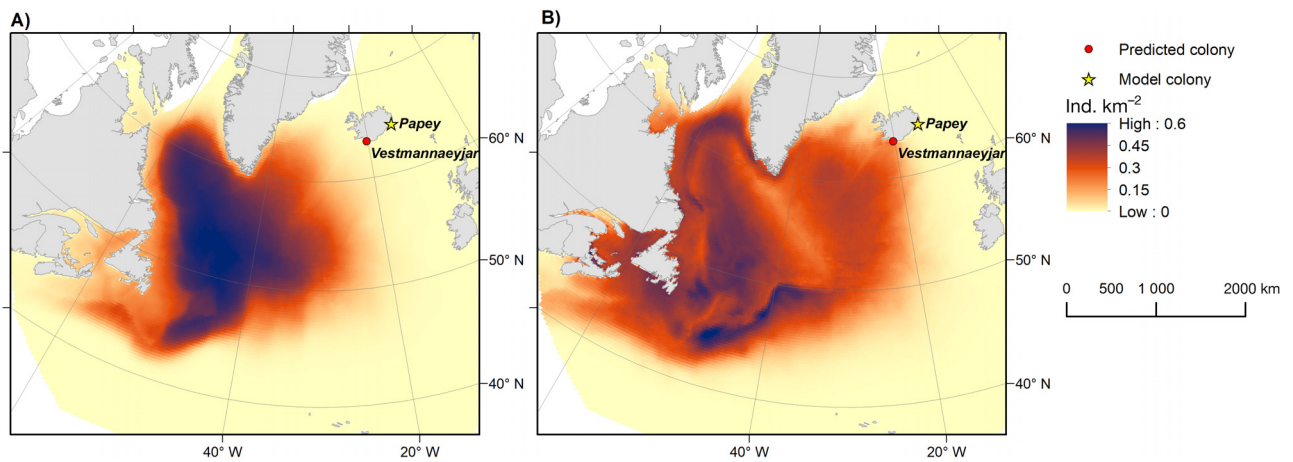
Fig. 4. (A) Predicted distribution and (B) 95% bootstrap CI range of Atlantic puffins from the breeding population at Vestmannaeyjar, Iceland, in December. Population size is 830 000 breeding pairs (1.66 million individuals; Hansen et al. 2011). Colour scale shows the density of birds predicted by the species distribution models on a linear scale from 0–0.6 birds km^{-2}
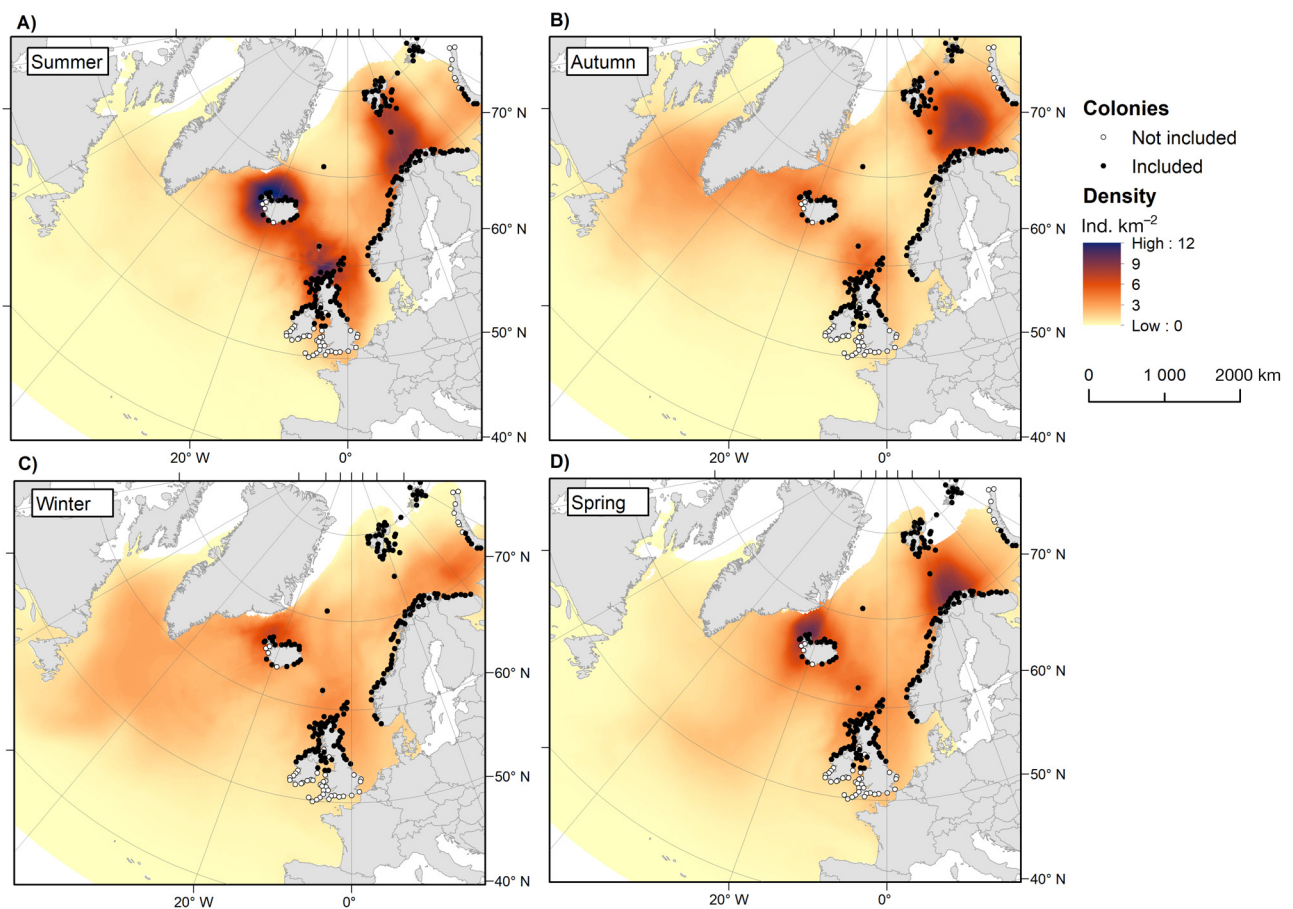

Fig. 5. Summed distribution of 6 Northeast Atlantic pelagic seabird species in (A) summer (May–July) breeding period, (B) autumn (August–October) migration period; (C) winter (November–January) wintering period and (D) spring (February–April) migration period. Black circles: colonies included in the density estimates; white circles: Northeast Atlantic colonies not included. The maps represent 23.5 million adult pelagic seabirds breeding in the Northeast Atlantic (86.6% of a total of 27.2 million birds). Colour scale shows the density of birds predicted by the species distribution models on a linear scale from 0–12 birds km^{-2}. The distribution of each species is shown in Figs. S4–S9

in the Denmark Strait and the southern Barents Sea, with a peak density of 9.9 birds km^{-2}. Compared to the breeding period, the pelagic seabirds of the Northeast Atlantic are more widely distributed during winter, i.e. they occur at lower densities over wider ocean areas. In this period, the pelagic seabirds are distributed across the ice-free part of the North Atlantic, from 50° N up to 75° N in the Barents Sea (Fig. 5C). Irrespective of season, the 3 areas holding the highest densities of birds were the Barents Sea, the Denmark Strait and the southern Norwegian Sea (i.e. the ocean areas around the Faroe Islands, Shetland and Orkney Islands).

3.4.2. Important ocean areas for Norwegian black-legged kittiwake populations

In the second example, we used the data set to identify important nonbreeding habitats for black-legged kittiwakes from the Norwegian mainland. Fig. 6 shows the estimated distribution with associated bootstrap CI range in 3 months (March, September and December). In March (Fig. 6A), before breeding, a large part of the population is concentrated in the southern Barents Sea. In September (Fig. 6B), the north-western Barents Sea is an important area. During winter, the kittiwake population migrates out of the Barents Sea and spreads out over
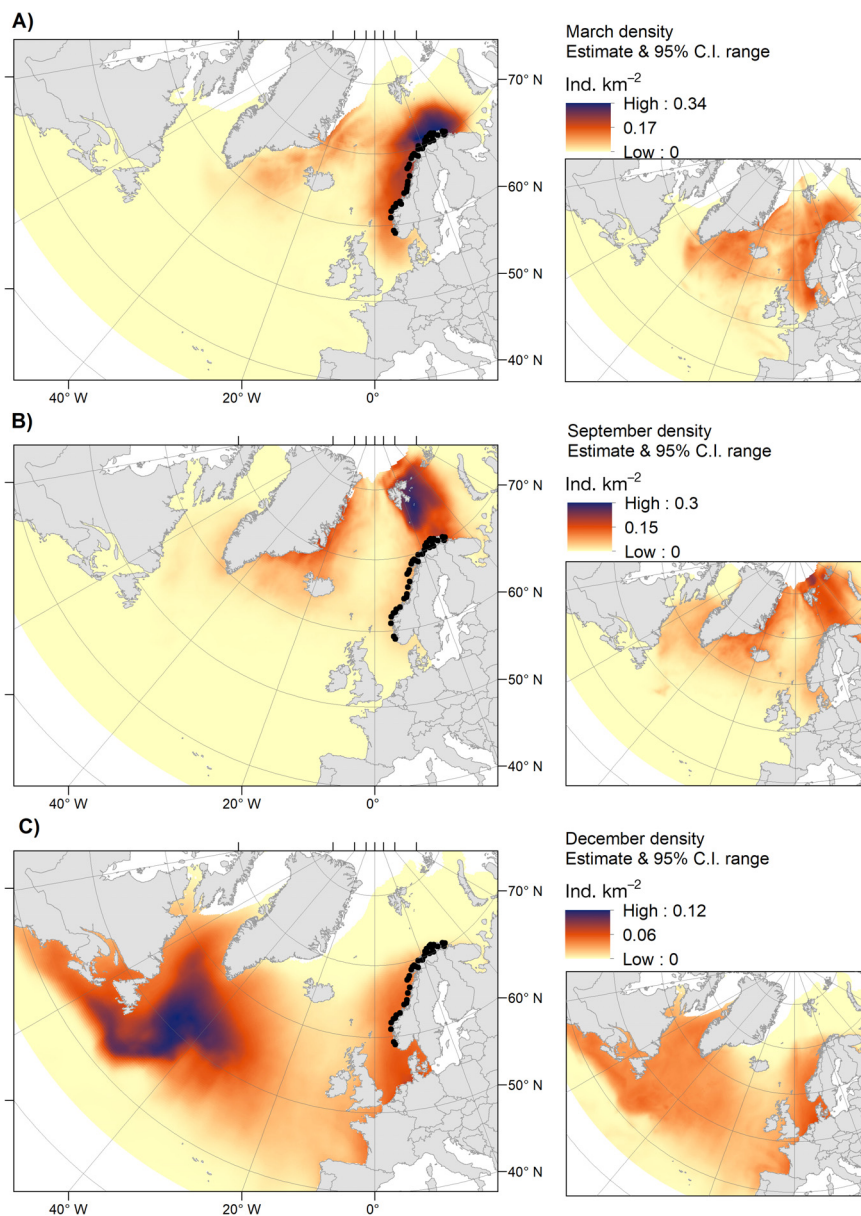
Fig. 6. Distribution of the Norwegian mainland populations of black-legged kittiwake in (A) March, (B) September and (C) December. Black circles: colonies included in the density estimates. Left panels: the density estimates; right panels: associated ranges in 95% bootstrap CIs. Maps represent 369 000 adult black-legged kittiwakes breeding on the Norwegian mainland (99.9% of a total population of 369 200 birds). Colour scale shows the density of birds predicted by the species distribution models on a linear scale. Note that the scales differ between (A), (B) and (C)

a large area in the Northwest Atlantic and the North Sea (Fig. 6C). The highest uncertainties in the estimates (Fig. 6, right panel) were found in the marginal areas of the habitat, e.g. along the coast of eastern Greenland in March and September and in the North Sea in December.

3.5. Spatial management applications

The data set can be used to identify the populations present within a given area and hence the populations impacted by a spatially restricted disturbance (e.g. accidental oil spill) or spatial management measures (e.g. MPA). In the following example, we used the data set to assess the populations of Northeast Atlantic pelagic seabirds that could be impacted by the recently proposed MPA in the mid-Atlantic (NACE). The estimated number of birds (including bootstrap estimates) present within the NACE area

were extracted for each colony and month. Four of the species in the data set (black-legged kittiwake, Atlantic puffin, northern fulmar and thick-billed murre) included colonies with more than 1% of the population within the NACE area in any month during the year (Fig. 7). In Fig. 8, we show detailed results including CI bands for 3 colonies with relatively high presence within the NACE area.

4. DISCUSSION

In this study, we developed a unique spatial data set of the predicted monthly distribution of 6 pelagic seabird species, covering 23.5 million adult birds, constituting 87% of their combined breeding populations in the Northeast Atlantic. The data set was based on year-round tracking data of adult seabirds from a network of seabird colonies, data describing the physical environment and data on population
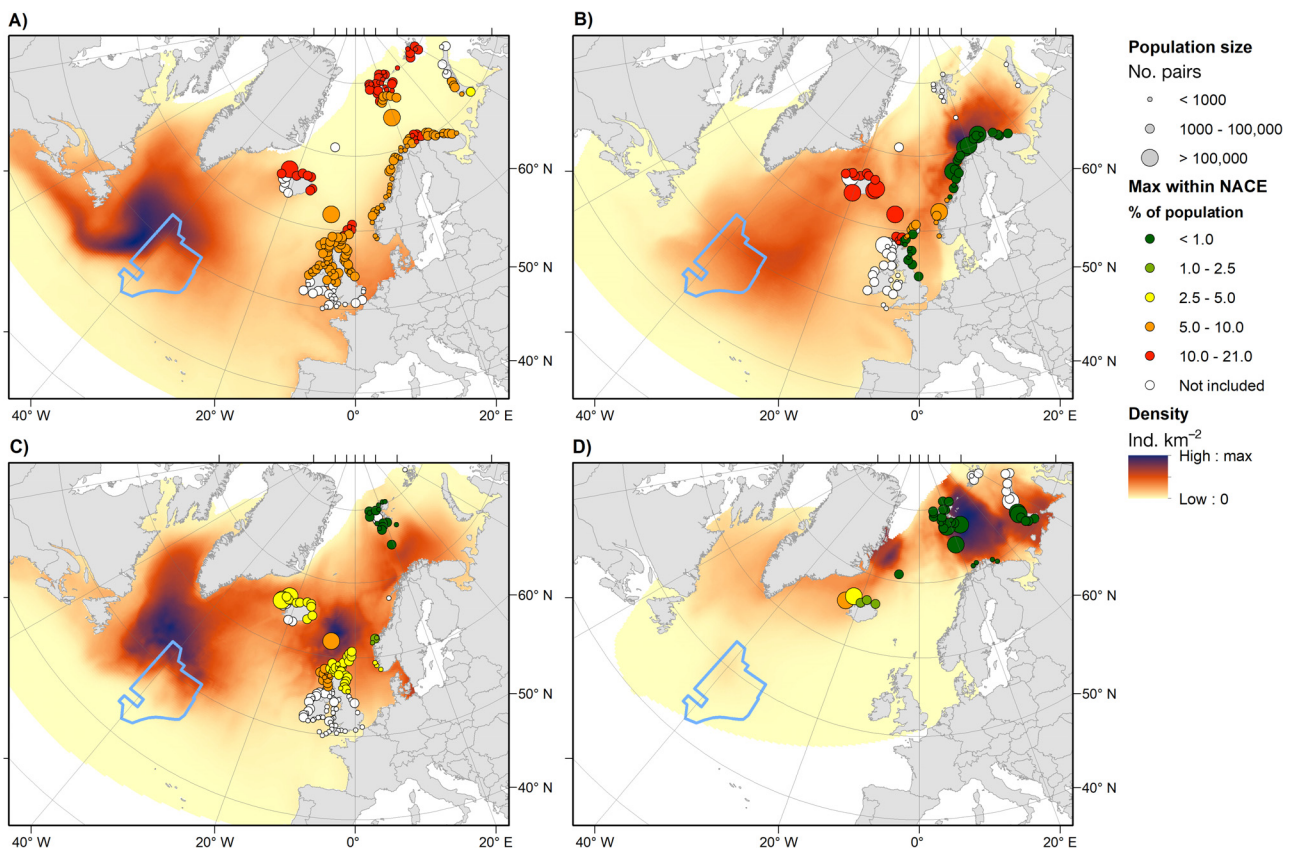

Fig. 7. Importance of the proposed marine protected area 'NACE' (light blue line) for Northeast Atlantic populations of (A) black-legged kittiwake, (B) Atlantic puffin, (C) northern fulmar and (D) thick-billed murre. Circles are seabird colonies with size representing colony size and colour representing the maximum annual proportion of the adult population found within the NACE area. Colour scale shows the total density of birds (number km^{-2}) predicted by the species distribution models on a linear scale for the month with the highest density of birds within the NACE area. The month with maximum number of birds within NACE were (A) December (black-legged kittiwake), (B) February (Atlantic puffin), (C) November (northern fulmar) and (D) October (thick-billed murre)
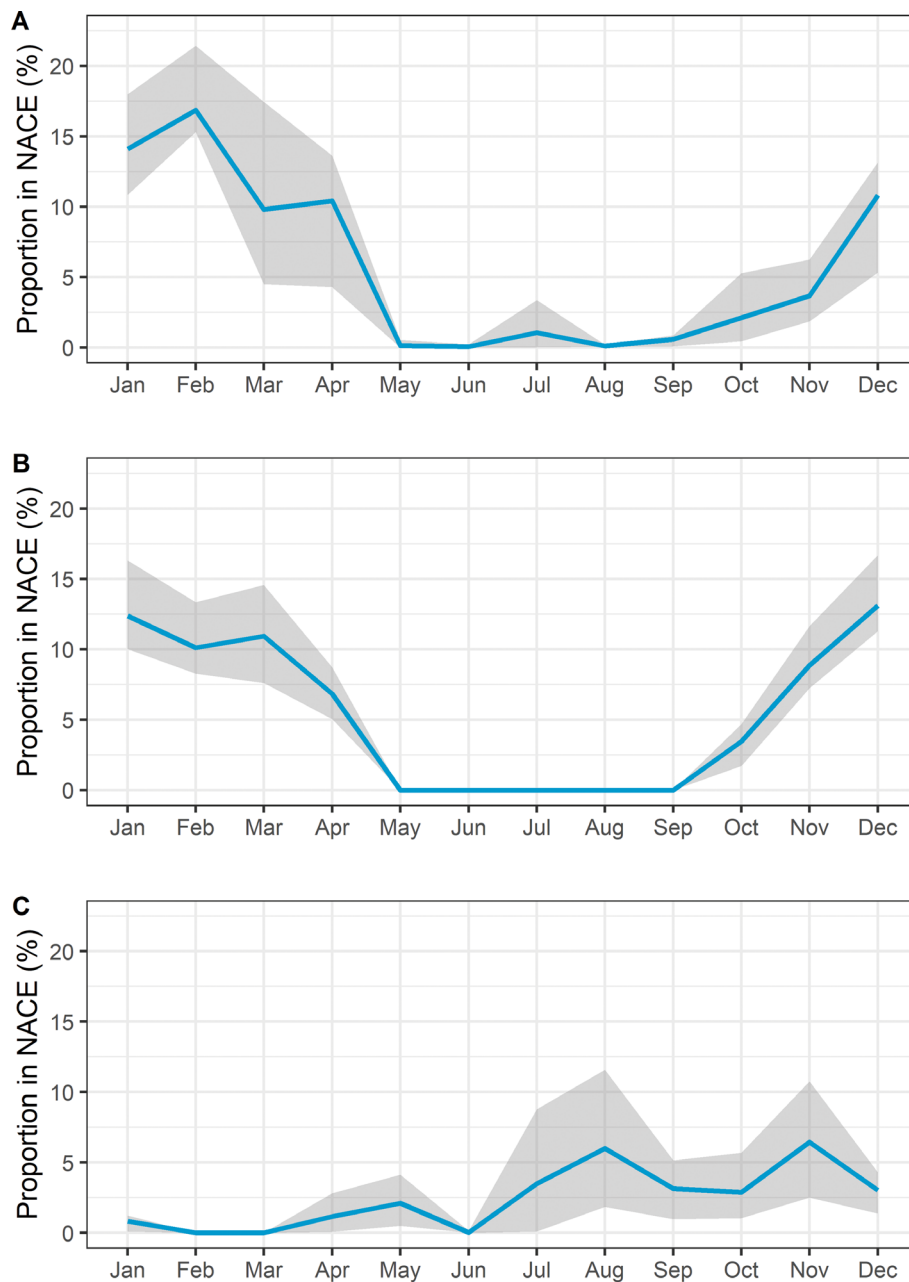
Fig. 8. Estimated proportion (±95% CI) of colony-specific populations within the proposed marine protected area (NACE) in the North Atlantic for (A) Atlantic puffins from Vestmannaeyjar, south Iceland (pop. size: 830 000 pairs), (B) black-legged kittiwakes from an unnamed colony at Nordaustlandet, Svalbard (pop. size: 4400 pairs) and (C) northern fulmars from the Faroe Islands (pop. size: 600 000 pairs). Confidence bands are calculated from cluster bootstrap samples

sizes from Ireland, the UK, Faroe Islands, Iceland, Norway and Russia. The resulting data set consisted of 4692 colony-specific rasters with a 0.25° × 0.25° pixel resolution giving the predicted number of birds per grid cell across the entire North Atlantic and associated uncertainty. The monthly temporal resolution, combined with colony-level predictions, make this data set a very powerful yet highly flexible tool

for use in a range of applications, from purely ecological/theoretical to more applied studies.

We demonstrated how the data set can be used in 2 very concrete management applications. First, in population management applications, the data set allows the identification of the seasonal habitats of a user-specified population. This information is especially relevant in assessments of critical habitats and

potential threats to endangered populations. Within the limits of the spatial extent of the colonies covered, the populations can be defined and aggregated on an optional level from the individual colony to marine ecosystems or administrative boundaries, and the bootstrap samples can be used to calculate CIs for the estimated spatial distribution. As an example, we summed the densities across all species and colonies in the data set to identify important seasonal ocean areas for Northeast Atlantic pelagic seabirds (Fig. 5). The Barents Sea, Denmark Strait and southern Norwegian Sea stand out as ocean areas with high densities of pelagic seabirds. In a second example, we identified the seasonal habitats of the Norwegian population of the black-legged kittiwake, which is listed as endangered on the Norwegian Red List (Henriksen & Hillmo 2015). Capelin *Mallotus villosus* is an important prey species for kittiwakes in the area (Reiertsen et al. 2014), and before the start of breeding season, high densities of kittiwakes were found in the southwestern Barents Sea coinciding with the capelin spawning migration (Fauchald & Erikstad 2002). Similarly, during autumn, kittiwakes were found in the north-western part of the Barents Sea, coinciding with the capelin feeding migration (Fauchald et al. 2006). During winter, the kittiwakes spread out in a large area in the Northwest Atlantic, possibly foraging on zooplankton prey (Frederiksen et al. 2012, Reiertsen et al. 2014).

In addition to population management applications, tracking data are increasingly used in spatial management applications (Lascelles et al. 2012, 2016). To this end, tracking data is typically used to assign biodiversity hotspots by identifying areas with a high occurrence of tracked birds (Le Corre et al. 2012, Montevecchi et al. 2012, Augé et al. 2018, Reisinger et al. 2018, Yurkowski et al. 2019, Carneiro et al. 2020, Cleasby et al. 2020, Hindell et al. 2020, Requena et al. 2020). This approach is, of course, sensitive to the species and populations represented by the tracking data, and in lieu of a valid link between populations and tracks, it is often assumed that a large number of tracked individuals, colonies and species will mitigate sampling biases. The data set developed in this study addresses this problem by specifically linking the habitat models to the populations they represent. Thus, the estimated number of birds within a specific area is corrected for population size. The data set can identify the origin of the birds in any given pixel, and importantly, clearly defines which populations are actually covered by the analysis. The data set can accordingly be used in assessments of MPAs, impact assessments related to

fisheries and offshore industrial development and risk assessments related to e.g. accidental oil spills. In the analysis of the proposed MPA in the North Atlantic (NACE) (Figs. 7 & 8), we could identify the species, colonies and seasons that would be affected by the proposed MPA. Importantly, the data set also identifies the populations not covered by the assessment. Clearly, the assessment is only valid for the part of the Northeast Atlantic populations of the 6 species covered by the current design. Regarding NACE, the assessment presented here is particularly limited as it does not cover seabirds from e.g. Greenland, North America, southern Ireland, southern UK and some colonies in eastern Iceland, as well as other species such as sooty shearwaters *Ardenna grisea* migrating from the South Atlantic (Hedd et al. 2012). This data gap illustrates the importance of validating the representativity of the tracking data with independent survey data and expert knowledge in applications related to marine spatial planning. Nevertheless, the present analysis demonstrates the importance of the NACE area for several important Northeast Atlantic populations of black-legged kittiwake, Atlantic puffin and northern fulmar.

With the increased use of tracking data to map mobile marine predators, the main mapping method has changed from an area-centred approach, where at-sea surveys were used to map the distribution in a given area, to a population-centred approach, where maps are based on the tracks of individuals from a population. As a consequence of this shift, the main sources of errors and bias have also changed. While at-sea surveys are prone to observer and detection bias and spurious relationships due to spatial autocorrelation along the transect lines (Redfern et al. 2006, Fauchald 2009), estimates of spatial patterns based on individual tracking data are susceptible to spurious relationships resulting from serial autocorrelation in individual space use (Fauchald & Tveraa 2006, Aarts et al. 2008). Accordingly, it is essential to provide robust model validations and estimates of the uncertainty around the model predictions. In the present study, cross-validation was used to address the validity and transferability of the models, and a cluster bootstrap procedure was used to construct CIs around the predictions. Positions provided by GLS have relatively low precision, and the errors depend on a variety of factors such as season, geographic location, species, weather and individual behaviour (Phillips et al. 2004, Lisovski et al. 2020). Accordingly, SDMs based on GLS can provide robust estimates of the coarse-scale (i.e. 100s of km) migration and distribution patterns (Phillips et al. 2004) but are less able to

discriminate fine-scale (i.e. 10s of km) habitat selection related to predictable environmental features such as e.g. frontal areas, seamounts or glacier fronts. Clearly, this limitation is also present in the data set developed here. Thus, the densities provided by the data set represent average densities over relatively large scales and do not account for the fine-scale clusters often observed in the distribution of seabirds at sea (Fauchald 2009). Finally, distribution models based on tracking data are highly susceptible to biases in the sample of tracked individuals (Aarts et al. 2008, Carneiro et al. 2020). The tracking data from the SEATRACK data set was limited to breeding individuals, excluding juveniles which constitute a large and important part of the population and potentially could have a different spatial distribution than adults (Carneiro et al. 2020). This limitation means that the data sets presented here are only valid for the adult, breeding component of the populations.

Equally important as a representative sampling of individuals (Carneiro et al. 2020) is the design with respect to the sampling of breeding colonies. Because different colonies might use different habitats, a study based on individuals from colonies that represent only a small portion of the population could give biased results. It is therefore essential to link the tracking data to the population that the data represent. In the present study, we used the network of SEATRACK colonies to assess the representativity of the tracking data. The analyses showed that birds from colonies close to each other (< 500 km) tended to use the same habitats, while for distant colonies (> 1000 km), the nonbreeding habitats were population-specific, and in many cases, non-overlapping. This result is corroborated by several North Atlantic multi-colony tracking studies demonstrating colony-specific and partly non-over-lapping winter habitats for black-legged kittiwakes (Frederiksen et al. 2012), common murres (Merkel et al. 2021a), thick-billed murres (Frederiksen et al. 2016, Merkel et al. 2021a), little auks (Fort et al. 2013) and Atlantic puffins (Fayet et al. 2017). The result has 2 important ramifications. First, it meant that we could use simple distance rules to link populations to tracking data and thus map population-specific distributions of birds. Second, and more broadly, the result has implications for the definition and use of predator hotspots in marine spatial management (Block et al. 2011, Lascelles et al. 2012). Predator hotspots, defined as areas where high abundances of species overlap in space and time (Davoren 2007), rely on the premise that spatial differences in biological production are reflected in the distribution of predators, and that high-productive areas, or hotspots, will attract preda-

tors from many populations and species. This bottom-up mechanism might, however, be weakened by predator–prey interactions (Fauchald 2009) as well as competition (Wakefield et al. 2013). When habitats are colony-specific, such as in our case, identification of hotspots based on individuals from a few locations could be problematic because important areas not covered by the sample will go undetected. The fine-meshed network of sampled colonies in the SEATRACK design allowed us to model a large portion of 6 pelagic seabird species from the Northeast Atlantic. Indeed, the pooled seabird distribution suggested that some areas (i.e. the Barents Sea, Denmark Strait and the southern Norwegian Sea) in general held higher concentrations of pelagic seabirds (Fig. 5). However, the analyses also showed that the birds spread out over large ocean areas after breeding, and by mid-winter the birds were widely distributed in the ice-free part of the North Atlantic, north of 50° N. Apparently, because different species and populations use different ocean areas, the hotspot approach should be used cautiously with respect to the pelagic seabirds in the Northeast Atlantic, as it could lead to overlooking vast regions with relatively lower densities that nonetheless represent large portion of the breeding populations.

Globally, seabird populations are in decline and threatened by a multitude of anthropogenic stressors (Dias et al. 2019). Threats in the marine environment include bycatch in fisheries, declining food availability due to overfishing and climate change, direct competition with fisheries, chronic and accidental oil pollution, littering and habitat loss to expanding marine industries and shipping (Dias et al. 2019). The increasing number of tracking studies, including the present one, provide new tools for identifying important seabird habitats and potential threats in the marine environment (Lascelles et al. 2012, Harrison et al. 2018, Hays et al. 2019). Importantly, this knowledge can be used to inform ecosystem-based marine spatial planning to alleviate the cumulative anthropogenic impacts on seabirds (Lascelles et al. 2012, Hays et al. 2019). To reduce the risk of information bias, the present study highlights the importance of large-scale coordinated efforts, like the SEATRACK initiative, to secure representative sampling and development of common data products. Moreover, as marine ecosystems are changing due to climate warming and ocean acidification, seabirds habitat and migration patterns will also change. To understand these changes and the impacts on seabirds, it is vital that such initiatives have a long-term perspective and that study design and data products are updated regularly.

LITERATURE CITED

Aarts G, MacKenzie M, McConnell B, Fedak M, Matthiopoulos J (2008) Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31:140–160

Amélineau F, Merkel B, Tarroux A, Descamps S and others (2021) Six pelagic seabird species of the North Atlantic engage in a fly-and-forage strategy during their migratory movements. Mar Ecol Prog Ser 676:127–144

Augé AA, Dias MP, Lascelles B, Baylis AMM and others (2018) Framework for mapping key areas for marine megafauna to inform marine spatial planning: the Falkland Islands case study. Mar Policy 92:61–72

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC Press, London

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: How, where and how many? Methods Ecol Evol 3:327–338

Barrett RT, Chapdelaine G, Anker-Nilssen T, Mosbech A and others (2006) Seabird numbers and prey consumption in the North Atlantic. ICES J Mar Sci 63:1145–1158

BirdLife International (2015) European Red List of Birds. http://datazone.birdlife.org/info/euroredlistsci

BirdLife International (2020) Species factsheet: *Rissa tridactyla*. http://datazone.birdlife.org/species/factsheet/black-legged-kittiwake-rissa-tridactyla

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ and others (2011) Tracking apex marine predator movements in a dynamic ocean. Nature 475:86–90

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Modell 157:281–300

Bråthen VS, Moe B, Amélineau F, Ekker M and others (2021) An automated procedure (v2.0) to obtain positions from light-level geolocators in large-scale tracking of

seabirds. A method description for the SEATRACK project. NINA Report 1893. Norwegian Institute for Nature Research, Trondheim

Broennimann O, Cola V Di, Guisan A (2020) ecospat: spatial ecology miscellaneous methods. R package version 3.1. https://cran.r-project.org/web/packages/ecospat/ecospat.pdf

Cadiou A, Riffaut L, McCoy KD, Cabelguen J and others (2004) Ecological impact of the 'Erika' oil spill: determination of the geographic origin of the affected common guillemots. Aquat Living Resour 17:369–377

Calenge C (2006) The package 'adehabitat' for the R software: a tool for the analysis of space and habitat use by animals. Ecol Modell 197:516–519

Carneiro APB, Pearmain EJ, Oppel S, Clay TA and others (2020) A framework for mapping the distribution of seabirds by integrating tracking, demography and phenology. J Appl Ecol 57:514–525

Cleasby IR, Owen E, Wilson L, Wakefield ED, O'Connell P, Bolton M (2020) Identifying important at-sea areas for seabirds using species distribution models and hotspot mapping. Biol Conserv 241:108375

Davies TE, Carneiro APB, Tarzia M, Wakefield E and others (2021) Multi-species tracking reveals a major seabird hotspot in the North Atlantic. Conserv Lett: e12824

Davison A, Hinkley D (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

Davoren GK (2007) Effects of gill-net fishing on marine birds in a biological hotspot in the Northwest Atlantic. Conserv Biol 21:1032–1045

Del Hoyo J, Elliot A, Sargatal J (1992) Handbook of the birds of the world. Lynx Editions, Barcelona

Delord K, Barbraud C, Bost CA, Deceuninck B and others (2014) Areas of importance for seabirds tracked from French southern territories, and recommendations for conservation. Mar Policy 48:1–13

Dias MP, Martin R, Pearmain EJ, Burfield IJ and others (2019) Threats to seabirds: a global assessment. Biol Conserv 237:525–537

Dormann CF, Elith J, Bacher S, Buchmann C and others (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC Press, Boca Raton, FL

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. Methods Ecol Evol 1:330–342

Fauchald P (2009) Spatial interaction between seabirds and prey: review and synthesis. Mar Ecol Prog Ser 391:139–151

Fauchald P, Erikstad KE (2002) Scale-dependent predator–prey interactions: the aggregative response of seabirds to prey under variable prey abundance and patchiness. Mar Ecol Prog Ser 231:279–291

Fauchald P, Tveraa T (2006) Hierarchical patch dynamics and animal movement pattern. Oecologia 149:383–395

Fauchald P, Mauritzen M, Gjøsæter H, Gjosaeter H (2006) Density-dependent migratory waves in the marine pelagic ecosystem. Ecology 87:2915–2924

Fauchald P, Tarroux A, Bråthen VS, Descamps S and others (2019) Arctic-breeding seabirds' hotspots in space and time: a methodological framework for year-round modelling of abundance and environmental niche using light-

logger data. NINA Report 1657. Norwegian Institute for Nature Research, Tromsø

Fayet al. Freeman R, Anker-Nilssen T, Diamond A and others (2017) Ocean-wide drivers of migration strategies and their influence on population breeding performance in a declining seabird. Curr Biol 27:3871–3878.e3

Fort J, Moe B, Strøm H, Grémillet D and others (2013) Multicolony tracking reveals potential threats to little auks wintering in the North Atlantic from marine pollution and shrinking sea ice cover. Divers Distrib 19:1322–1332

Frederiksen M (2010) Appendix 1: seabirds in the North East Atlantic. A review of status, trends and anthropogenic impact. In: Action plan for seabirds in Western-Nordic areas. Nordic Council of Ministers, Copenhagen, p 47–122

Frederiksen M, Moe BB, Daunt F, Phillips RA and others (2012) Multicolony tracking reveals the winter distribution of a pelagic seabird on an ocean basin scale. Divers Distrib 18:530–542

Frederiksen M, Descamps S, Erikstad KE, Gaston AJ and others (2016) Migration and wintering of a declining seabird, the thick-billed murre *Uria lomvia*, on an ocean basin scale: conservation implications. Biol Conserv 200: 26–35

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Modell 135:147–186

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8: 993–1009

Hansen ES, Sigursteinsson M, Gardarsson A (2011) The breeding population size of Atlantic puffin in Vestmannaeyjar, S-Iceland. Bliki 31:15–24

Harrison AL, Costa DP, Winship AJ, Benson SR and others (2018) The political biogeography of migratory marine predators. Nat Ecol Evol 2:1571–1578

Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall/CRC Press, Boca Raton, FL

Hays GC, Bailey H, Bograd SJ, Bowen WD and others (2019) Translating marine animal tracking data into conservation policy and management. Trends Ecol Evol 34: 459–473

Hedd A, Montevecchi WA, Otley H, Phillips RA, Fifield DA (2012) Trans-equatorial migration and habitat use by sooty shearwaters *Puffinus griseus* from the South Atlantic during the nonbreeding season. Mar Ecol Prog Ser 449: 277–290

Henriksen S, Hillmo O (2015) Norsk rødliste for arter 2015. Artsdatabanken, Trondheim

Hijmans RJ (2020) raster: geographic data analysis and modeling. R package version 3.4-5. https://CRAN.R-project.org/package=raster

Hindell MA, Reisinger RR, Ropert-Coudert Y, Hückstädt LA and others (2020) Tracking of marine predators to protect Southern Ocean ecosystems. Nature 580:87–92

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Modell 199:142–152

Hyrenbach KD, Veit RR, Weimerskirch H, Hunt GL (2006) Seabird associations with mesoscale eddies: the subtropical Indian Ocean. Mar Ecol Prog Ser 324:271–279

Lascelles BG, Langham GM, Ronconi RA, Reid JB (2012) From hotspots to site protection: identifying marine protected areas for seabirds around the globe. Biol Conserv 156:5–14

Lascelles BG, Taylor PR, Miller MGR, Dias MP and others

(2016) Applying global criteria to tracking data to define important areas for marine conservation. Divers Distrib 22:422–431

Le Corre M, Jaeger A, Pinet P, Kappes MA and others (2012) Tracking seabirds to identify potential marine protected areas in the tropical western Indian Ocean. Biol Conserv 156:83–93

Léandri-Breton DJ, Tarroux A, Elliott K, Legagneux P and others (2021) Long-term tracking of an Arctic-breeding seabird indicates high fidelity to pelagic wintering areas. Mar Ecol Prog Ser 676:205–218

Lisovski S, Bauer S, Briedis M, Davidson SC and others (2020) Light-level geolocator analyses: a user's guide. J Anim Ecol 89:221–236

Lobo JM, Jiménez-valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151

Melo-Merino SM, Reyes-Bonilla H, Lira-Noriega A (2020) Ecological niche models and species distribution models in marine environments: a literature review and spatial analysis of evidence. Ecol Modell 415:108837

Merkel B, Descamps S, Yoccoz NG, Grémillet D and others (2021a) Individual migration strategy fidelity but no habitat specialization in two congeneric seabirds. J Biogeogr 48:263–275

Merkel B, Descamps S, Yoccoz NG, Grémillet D and others (2021b) Strong migratory connectivity across metapopulations of sympatric North Atlantic seabirds. Mar Ecol Prog Ser 676:173–188

Montevecchi WA, Hedd A, McFarlane Tranquilla L, Fifield DA and others (2012) Tracking seabirds to identify ecologically important and high risk marine areas in the western North Atlantic. Biol Conserv 156:62–71

Norberg A, Abrego N, Blanchet FG, Adler FR and others (2019) A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecol Monogr 89:1370

OSPAR (2021) The 'North Atlantic Current and Evlanov Seamount MPA'. https://www.ospar.org/work-areas/bdc/marine-protected-areas/ospar-seeks-views-on-the-nomination-proforma-for-the-north-atlantic-current-and-evlanov-seamount-mpa (accessed 1 Aug 2021)

Péron C, Authier M, Grémillet D (2018) Testing the transferability of track-based habitat models for sound marine spatial planning. Divers Distrib 24:1772–1787

Perrow MR, Harwood AJP, Skeate ER, Praca E, Eglington SM (2015) Use of multiple data sources and analytical approaches to derive a marine protected area for a breeding seabird. Biol Conserv 191:729–738

Phillips RA, Silk JRD, Croxall JP, Afanasyev V, Briggs DR (2004) Accuracy of geolocation estimates for flying seabirds. Mar Ecol Prog Ser 266:265–272

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19:181–197

R Development Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Redfern JV, Ferguson MC, Becker EA, Hyrenbach KD and others (2006) Techniques for cetacean-habitat modeling. Mar Ecol Prog Ser 310:271–295

Reiertsen TK, Erikstad KE, Anker-Nilssen T, Barrett RT and others (2014) Prey density in non-breeding areas affects adult survival of black-legged kittiwakes *Rissa tridactyla*

breeding in the southern Barents Sea. Mar Ecol Prog Ser 509:289–302

Reisinger RR, Raymond B, Hindell MA, Bester MN and others (2018) Habitat modelling of tracking data from multiple marine predators identifies important areas in the Southern Indian Ocean. Divers Distrib 24:535–550

Requena S, Oppel S, Bond AL, Hall J and others (2020) Marine hotspots of activity inform protection of a threatened community of pelagic species in a large oceanic jurisdiction. Anim Conserv 23:585–596

Sansom A, Wilson LJ, Caldow RWG, Bolton M (2018) Comparing marine distribution maps for seabirds during the breeding season derived from different survey and analysis methods. PLOS ONE 13:e0201797

Scales KL, Miller PI, Ingram SN, Hazen EL, Bograd SJ, Phillips RA (2016) Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. Divers Distrib 22:212–224

Strøm H, Descamps S, Ekker M, Fauchald P, Moe B (2021) Tracking the movements of North Atlantic seabirds: steps towards better understanding of population dynam-

ics and marine ecosystem conservation. Mar Ecol Prog Ser 676:97–116

Technitis G, Othman W, Safi K, Weibel R (2015) From A to B, randomly: a point-to-point random trajectory generator for animal movement. Int J Geogr Inf Sci 29:912–934

Tremblay Y, Bertrand S, Henry RW, Kappes MA, Costa DP, Shaffer SA (2009) Analytical approaches to investigating seabird-environment interactions: a review. Mar Ecol Prog Ser 391:153–163

Wakefield ED, Bodey TW, Bearhop S, Blackburn J and others (2013) Space partitioning without territoriality in gannets. Science 341:68–70

Wisz MS, Hijmans RJ, Li J, Peterson AT and others (2008) Effects of sample size on the performance of species distribution models. Divers Distrib 14:763–773

Wood SN (2017) Generalized additive models: an introduction with R, 2nd edn. CRC Press, Boca Raton, FL

Yurkowski DJ, Auger-Méthé M, Mallory ML, Wong SNP and others (2019) Abundance and species diversity hotspots of tracked marine predators across the North American Arctic. Divers Distrib 25:328–345

Appendix.

Per Fauchald[1],*, Arnaud Tarroux[1], Françoise Amélineau[2], Vegard Sandøy Bråthen[3], Sébastien Descamps[2], Morten Ekker[4], Halfdan Helgi Helgason[2,21], Malin Kjellstadli Johansen[2], Benjamin Merkel[2,5], Børge Moe[3], Jens Åström[3], Tycho Anker-Nilssen[3], Oskar Bjørnstad[6], Olivier Chastel[7], Signe Christensen-Dalsgaard[3], Jóhannis Danielsen[8], Francis Daunt[9], Nina Dehnhard[3], Kjell Einar Erikstad[1], Alexey Ezhov[10], Maria Gavrilo[11,12,22], Gunnar Thor Hallgrimsson[13], Erpur Snær Hansen[14], Mike Harris[9], Morten Helberg[15], Jón Einar Jónsson[16], Yann Kolbeinsson[17], Yuri Krasnov[10], Magdalene Langset[3], Svein-Håkon Lorentsen[3], Erlend Lorentzen[2], Mark Newell[9], Bergur Olsen[8], Tone Kristin Reiertsen[1], Geir Helge Systad[18], Paul Thompson[19], Thorkell Lindberg Thórarinsson[17], Sarah Wanless[9], Katarzyna Wojczulanis-Jakubas[20], Hallvard Strøm[2]

[1]Norwegian Institute for Nature Research, Fram Centre, 9296 Tromsø, Norway
[2]Norwegian Polar Institute, Fram Centre, 9296 Tromsø, Norway
[3]Norwegian Institute for Nature Research, 7485 Trondheim, Norway
[4]The Norwegian Environment Agency, 7485 Trondheim, Norway
[5]Akvaplan-niva AS, Fram Centre, 9296 Tromsø, Norway
[6]Karmøy Ringmerkingsgruppe, 4280 Skudeneshavn, Norway
[7]Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 CNRS-Univ. La Rochelle, France
[8]Faroe Marine Research Institute, 110 Tórshavn, Faroe Islands
[9]UK Centre for Ecology & Hydrology, Bush Estate, Penicuik EH26 0QB, UK
[10]Murmansk Marine Biological Institute of the Russian Academy of Sciences (RAS), Murmansk 183010, Russia
[11]National Park Russian Arctic, 163000 Arkhangelsk, Russia
[12]Association Maritime Heritage, Icebreaker 'Krassin', Saint-Petersburg 199106, Russia
[13]Faculty of Life and Environmental Sciences, University of Iceland, 102 Reykjavik, Iceland
[14]South Iceland Nature Research Centre, 900 Vestmannaeyjar, Iceland
[15]Østfold University College, 1757 Halden, Norway
[16]University of Iceland's Research Center at Snæfellsnes, 340 Stykkishólmur, Iceland
[17]Northeast Iceland Nature Research Centre, 640 Húsavík, Iceland
[18]Norwegian Institute for Nature Research, 5006 Bergen, Norway
[19]Lighthouse Field Station, School of Biological Sciences, University of Aberdeen, IV11 8LU, UK
[20]University of Gdansk, Dept. of Vertebrate Ecology and Zoology, 80-308 Gdansk, Poland

[21]*Present address:* East Iceland Nature Research Centre, 740 Neskaupstaður, Iceland

[22]*Present address:* Arctic and Antarctic Research Institute, Saint-Petersburg 199397, Russia