ECOGRAPHY

Research

A methodological framework to predict the individual and population-level distributions from tracking data

Philippine Chambault, Tarek Hattab, Pascal Mouquet, Touria Bajjouk, Claire Jean, Katia Ballorain, Stéphane Ciccione, Mayeul Dalleau and Jérôme Bourjea

P. Chambault (https://orcid.org/0000-0003-4054-6334) ✉ (philippine.chambault@gmail.com), T. Hattab and J. Bourjea, MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Sète, France, and: IFREMER Inst. Français pour l'Exploitation de la Mer, Sète, France. – P. Mouquet, Saint Leu, La Réunion, France. – T. Bajjouk, Inst. Français pour l'Exploitation de la Mer (IFRMER), DYNECO/LEBCO, Plouzané, France. – C. Jean and S. Ciccione, Kelonia, l'Observatoire des Tortues Marines, Saint Leu, La Réunion, France. – K. Ballorain and M. Dalleau, Centre d'Etude et de Découverte des Tortues Marines (CEDTM), Saint-Leu, La Réunion, France. KB, Biodiversity French Agency, Mayotte and Glorieuses Marine Nature Parks, Saint-Leu, La Réunion, France.

Ecography
44: 1–12, 2021
doi: 10.1111/ecog.05436

Subject Editor: Nigel G. Yoccoz
Editor-in-Chief:
Jens-Christian Svenning
Accepted 10 January 2021

Despite the large number of species distribution modelling (SDM) applications driven by tracking data, individual information is most of the time neglected and traditional SDM approaches commonly focus on predicting the potential distribution at the species or population-level. By running classical SDMs (population approach) with mixed models including a random factor to account for the variability attributable to individual (individual approach), we propose an innovative five-steps framework to predict the potential and individual-level distributions of mobile species using GPS data collected from green turtles. Pseudo-absences were randomly generated following an environmentally-stratified procedure. A negative exponential dispersal kernel was incorporated into the individual model to account for spatial fidelity, while five environmental variables derived from high-resolution Lidar and hyperspectral data were used as predictors of the species distribution in generalized linear models. Both approaches showed a strong predictive power (mean: AUC > 0.93, CBI > 0.88) and goodness-of-fit (0.6 < adjusted R^2 < 0.9), but differed geographically with favorable habitats restricted around the tagging locations for the individual approach whereas favorable habitats from the population approach were more widespread. Our innovative way to combine predictions from both approaches into a single map provides a unique scientific baseline to support conservation planning and management of many taxa. Our framework is easy to implement and brings new opportunities to exploit existing tracking dataset, while addressing key ecological questions such as inter-individual plasticity and social interactions.

Keywords: GPS tracking, green turtles, Indian Ocean, pseudo-absences, Shannon index, spatial modelling

NORDIC SOCIETY OIKOS

www.ecography.org

© 2021 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

Over the past two decades, the need to quantify species distribution inspired the development of various mechanistic (i.e. process-based) and correlative tools known as species distribution models (SDMs) which are now available to ecologists interested in predicting species distribution. SDMs are of particular interest when dealing with endangered and invasive species as they support effective conservation planning, e.g. identification of potential protected areas (Krüger et al. 2017, Domisch et al. 2019) or areas susceptible to biological invasions (Hattab et al. 2017) or climate change (Erauskin-Extramiana et al. 2019).

Correlative species distribution models (SDMs), grounded in ecological niche theory (Hutchinson 1957), are statistical tools commonly used to predict suitable habitats of a species based on the statistical relationship between its occurrence and its environment (Austin 2002, Elith and Leathwick 2009). In contrast to correlative models, mechanistic SDMs use physiological information about a species to determine the range of environmental conditions within which the species can persist (Kearney and Porter 2009). Several studies demonstrated the benefits of mechanistic approaches to predict spatiotemporal variation in the distribution and abundance of species (Gaspar and Lalire 2017, Lalire and Gaspar 2019, Gaspar et al. 2020, Putman et al. 2020). However, the use of simpler and more popular correlative SDMs presents a significant advantage when the lack of knowledge on the species physiological tolerance (i.e. optimum and range) is counterbalanced by the availability of relevant environmental variables over the species geographical range. Accordingly, correlative SDMs have been more largely used across terrestrial, freshwater and marine realms compared to mechanistic models.

Correlative SDMs commonly predict potentially suitable environmental conditions for a species. In conservation spatial planning, the potential distribution of a species is a powerful information tool to delineate protected areas in a more efficient way. Although discerning potentially suitable areas for a given species is an important asset in conservation, it is also necessary to estimate current distributions (i.e. the realized distribution: area actually occupied by the species) in order to conserve current populations (Marcer et al. 2013). When using correlative species distribution modelling techniques in conservation assessment, these differences between realized and potential distributions need to be explicitly accounted for and explained (Marcer et al. 2013). However, the number of studies providing a unified modelling framework that combines the realized and potential distributions is very limited (Hattab et al. 2017).

Traditionally, correlative SDMs have viewed the niche as a property of the species or population as a whole, treating conspecific individuals as ecologically equivalent. Many apparently generalized species are in fact composed of individual specialists that use small subsets of the population's niche (Bolnick et al. 2002). The degree of individual specialization varies widely among species and among populations, reflecting a diverse array of physiological, behavioral and ecological mechanisms that can generate intra-population variation (Bolnick et al. 2003). For mobile species, the inter-individual variation in spatial distributions is not only the result of the phylogenetic history, but it could also be induced by a plastic character related to decision-making (Cassini 2013). For such species, habitat selection is an individual process and a behavioral phenomenon that determines the realized distribution (i.e. habitat use within the home range). Habitat selection implies the choice of an alternative among different behavioral options available and the result of these behaviors is an individual pattern of habitat use (Cassini 2013). The population's realized distribution is thus the sum or result of several spatial behaviors that are organized in a hierarchy starting from the individual level.

SDMs have been criticized for assuming that different subpopulations from the same species will respond similarly to climate-induced perturbations, therefore avoiding potential intraspecific differentiations occurring across the species range. To overcome this issue, several studies have recently incorporated information on biotic interactions into SDMs (e.g. through phylogeographic structure) and showed how genetic and environmental variation within species ranges can affect SDM predictions (Hällfors et al. 2016, Lecocq et al. 2019, Chardon et al. 2020). For example, a recent study conducted on the arctic–alpine cushion plant constructed intraspecific-level SDMs and showed that both genetic and habitat-informed SDMs were considerably more accurate than a classical species-level model (Chardon et al. 2020), reinforcing the need of comparing population-based models to whole-species models.

Numerous correlative SDMs based on animal tracking data have been developed in recent years for a broad range of taxa in both terrestrial (Wisz et al. 2008, Marini et al. 2010, Alamgir et al. 2015) and marine realms (Varo-Cruz et al. 2016, Scales et al. 2017, Brodie et al. 2018, Abrahms et al. 2019). However, the identity of tracked individuals is never included into SDMs to predict the realized distribution of a species. Inter-individual heterogeneity is commonly observed in many taxa but it is rarely taken into account since traditional SDM approaches commonly neglect the individual-level and by fitting population-level SDMs using pooled occurrence records. To provide a full picture of the potential (area that could be occupied) and realized (area actually occupied) distributions of mobile species, we propose here a 5-step species distribution modelling framework based on tracking data. This framework considers a classical SDM for population-level predictions and mixed-models for individual level inferences.

To apply our framework on a real case study, we used GPS occurrences from 21 juvenile green turtles satellite tracked in Reunion Island waters in the south-west Indian Ocean. Inter-individual plasticity in movements is commonly observed in sea turtles and particularly in the green turtle (Hays et al. 2002, Schofield et al. 2010, Dalleau et al. 2014, Luschi et al. 2017, Dujon et al. 2018). A recent study conducted on 49 juvenile green turtles tracked from five foraging grounds in

the Indian Ocean has shown contrasting behaviors from individuals inhabiting the same site (Chambault et al. 2020a). Therefore, this species represents an excellent case study to test our framework. To produce predictions that reflect both the potential and the realized niches, GPS occurrences were linked to five environmental predictors derived from high resolution Lidar and hyperspectral data (Bajjouk et al. 2019) using SDMs (potential distribution) and mixed-models (realized distribution). Finally, we produced a useful map of the combined approaches, representing the two facets of the population's spatial distribution (i.e. potential versus realized) that would undoubtedly be a useful tool to address key questions in biogeography, behavioural ecology and conservation planning.

Material and methods

Study area and tag deployment

The study area spreads from 21°38' to 20°87'S and from 55°21' to 55°83'E, and is located in a French overseas territory of the south-west Indian Ocean: La Reunion (Fig. 1). Only a small portion of the total geographical range of this specie was used due to the strong site fidelity already evidenced in La Reunion (Chambault et al. 2020a) for immature green turtles. Between 2010 and 2019, 21 juvenile green turtles were caught and satellite tagged in La Reunion from three different sectors located on the west coast (Fig. 1).

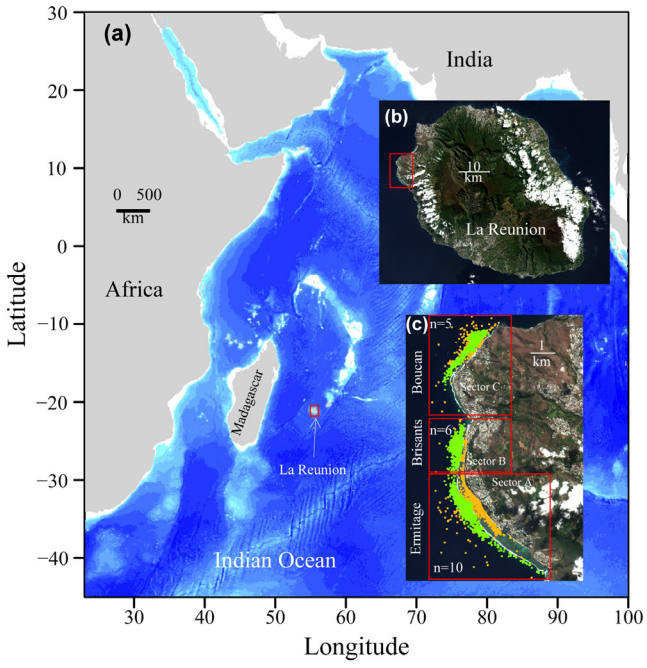

Figure 1. Study area of La Reunion (red rectangle in (a)), (b) refers to La Reunion, (c) to the three tagging sites (Boucan, Brisants and Ermitage) with the GPS locations of the turtles on the reef flat (in orange) and beyond isobaths −30 m and on the slope habitat (in green and used for predictions background).

The three sectors were chosen based on turtles sightings (Jean et al. 2010) and on their contrasting habitats (J. B. Nicet pers. comm.):

- Sector A (Ermitage, n = 10): located south of the study area and characterized by a large fore reef and reef flat and a blind reef pass.
- Sector B (Brisants, n = 5): located north of sector A and characterized by a fore reef and reef flat of medium width.
- Sector C (Boucan, n = 6): located at the north of the study area and characterized by a narrow fore reef and reef flat.

In-water turtles were captured by scuba diving. Once captured, the curved carapace length (CCL) was measured following Eckert et al.'s procedure (1999), and body mass was taken using an electronic dynamometer. Argos Fastloc-GPS tags (Wildlife Computers Redmond, WA, USA) that provide Fastloc-GPS data relayed via the Argos satellite system (<www.argos-system.org/>) were then fixed on each juvenile green turtle (Chambault et al. 2020a for details). In order to increase the number of positions recorded, the tags were programmed to record GPS locations at a sampling interval set at 30 min.

Data pre-filtering

Due to the restricted dispersal pattern commonly observed in juvenile green turtles in their coastal habitats and the large uncertainties associated with Argos locations, only Fastloc-GPS locations were retained for the analysis to improve the quality of the results and provide reliable kernel estimates (Thomson et al. 2017). The Fastloc-GPS data were filtered to reduce measurement errors by removing locations with residuals values above 35 and locations recorded by less than five satellites (Dujon et al. 2014). We restricted our dataset to positions associated with a travel speed lower than 5 km h^{-1} (Schofield et al. 2010). Finally, remaining positions located on land were discarded. The tracking data are summarized in the Supporting information.

Chambault et al (2020a) have shown a diel pattern in terms of habitat use between the diurnal *reef flat* habitat and the nocturnal *slope* habitat for green turtles of La Reunion. We based our hypothesis on visual sightings of turtles in this area showing the use of steep, rough and concave areas as foraging and resting habitats during the day. We therefore expect to that green turtles will target steep, rough and concave areas. That is why we decided to focus on the diurnal habitat *slope* in the reef front zone (from 3 to 40 m deep) as the geographic extent to run the analysis.

Environmental variables

Five physical variables were used to investigate the drivers of the turtles' coastal movements and predict their distribution. These variables were derived from high resolution Lidar (Light detection and ranging) and hyperspectral airborne data collected in 2009 and 2010, and processed over the west coast of La Reunion (Mouquet et al. 2015a, Bajjouk et al.

2019). Processing of these data provided the following layers (Supporting information):

1) *The bathymetry.* The fine-scale gridded bathymetry (spatial resolution of 1 m, down to a maximum depth of 40 m) was provided by the SHOM (Service Hydrographique et Océanographique de la Marine, <https://data.shom.fr/donnees>) acquired with a bathymetric Lidar sensor.
2) *The slope.* The maximum rate of change in elevation between a given point and its surrounding points (expressed in degrees) was calculated from Lidar in 3 × 3 m grid cell, and then the median was extracted in cells of 19 × 19 m.
3) *The slope variance.* This variable was used as a proxy of the bottom roughness. The slope was first calculated in 3 × 3 m grid cells. The variance was then calculated in 9 × 9 m cells (expressed in degrees^2), and the median of the variance was finally derived in 19 × 19 m cells.
4) *The concavity.* This variable was used as a proxy for potential resting areas used by the turtles, e.g. caves. The profile convexity (e.g. vertical component of curvature, expressed in degrees m^{-1}) was first calculated on a 9 × 9 m grid to detect convex and concave profiles. The median in a 9 × 9 m grid was then calculated to remove measurement artefacts, and only the negative values were retained as associated with concave profiles (i.e. hollows). The median opposite values of concavity were finally calculated in a 19 × 19 m grid to extend the range of influence.
5) *The distance to rocks.* This variable was used to identify potential feeding areas (e.g. benthic algae on hard bottom) and/or resting areas (e.g. caves). The hard (rocks and reef patches) versus soft substrates (sand and rubbles) were first discriminated based on simple thresholding of the image data 'Pseudo-colour images of the seabed of the reef areas of the west coast of Reunion Island' (Mouquet et al. 2015a) from the HYSCORES project (Mouquet et al. 2015b). This data was extracted from hyperspectral images at a 40 cm resolution. The shortest distance between each grid cell and the closest identified rock pixel was then calculated in a grid of 1 × 1 m.

All environmental predictors were then resampled to get the same spatial resolution of 5 × 5 m.

Species distribution modelling

In order to model the different aspects of the geographical distribution of juvenile green turtles, we built two types of SDMs. The first SDM (called hereafter 'Population approach') reflecting the potential distribution of the entire population was fitted using generalized linear models (GLMs) applied to the occurrences' records of the 21 individuals and a pseudo-absences dataset. The second model (called hereafter 'Individual approach') inferred the realized distributions of each of the 21 individuals and reflects the portion of the potential niche space that is effectively occupied by individuals. Generalized linear mixed-models (GLMMs, from the *lme4* package in R) including the *individual* as random factor were fitted by incorporating dispersal-related covariate (Dispersal coefficient) in the model in addition to the five environmental variables. The five-steps methodology is described in Fig. 2 as follows.

Pseudo-absences data generation

We used an environmental background based technique to generate pseudo-absences (Senay et al. 2013, Iturbide et al. 2015, Hattab et al. 2017, Ben Rais Lasram et al. 2020, Schickele et al. 2020), relying on the assumption that true absences are more likely located in areas that are environmentally dissimilar from presence locations. A principal component analysis (PCA) was used to generate a two-dimensional environmental background (Fig. 2, step 1) representing the ordination results of the five environmental variables. Occurrences records were projected on this environmental background to delineate environmental combinations that were suitable to the turtles. The suitable environmental background was considered as the smallest convex hyper-volume in the environmental space containing species observation records. A restricted convex hull excluding occurrence points within the 2.5% and 97.5% percentiles for each ordination axis has been defined (i.e. excluding observations in the most extreme environmental conditions). As recommended by the 'D-designs' theory (Montgomery 2007), pseudo-absences were then randomly generated outside this restricted convex hull in equal number to the filtered occurrences. Finally, pseudo-absences were projected back in geographical cells showing environmental conditions outside species' environmentally favorable areas (Supporting information).

For the population approach, the convex hull was common to all individuals (Fig. 2, step 1 right) while for the Individual approach, one convex hull was generated for each individual based on the occurrences of each turtle (Fig. 2, step 1 left). To increase the robustness of the results and assess their sensitivity to the pseudo-absences generation procedure, 10 different sets of pseudo-absences were simulated (i.e. 10 runs for the population approach and 10 runs for each of the 21 individuals for the individual approach).

Dispersal-related variable calculation

To take into account the spatial fidelity of each turtle in the Individual approach, a dispersal-related variable was included in the GLMMs. We used a negative exponential dispersal kernel (Meentemeyer et al. 2008) to quantify the degree of fidelity of each individual to each pixel of 5 m within the study area (hereafter called 'Dispersal coefficient': *DC*) – Fig. 2, step 2 left.

The parameter *DC* modifies the form of the dispersal kernel where low values of *DC* indicate localized fidelity in the vicinity of occupied sites, whereas high values indicate potential for dispersal from any occupied site. This *DC* was calculated using the *iForce* function from the *iSDM* package (Hattab and Lenoir 2016). A logarithmic sequence of 10 *DC* values ranging between 0 and 1 was tested for each turtle to set the optimal *DC* value ('Individual approach via GLMMs'

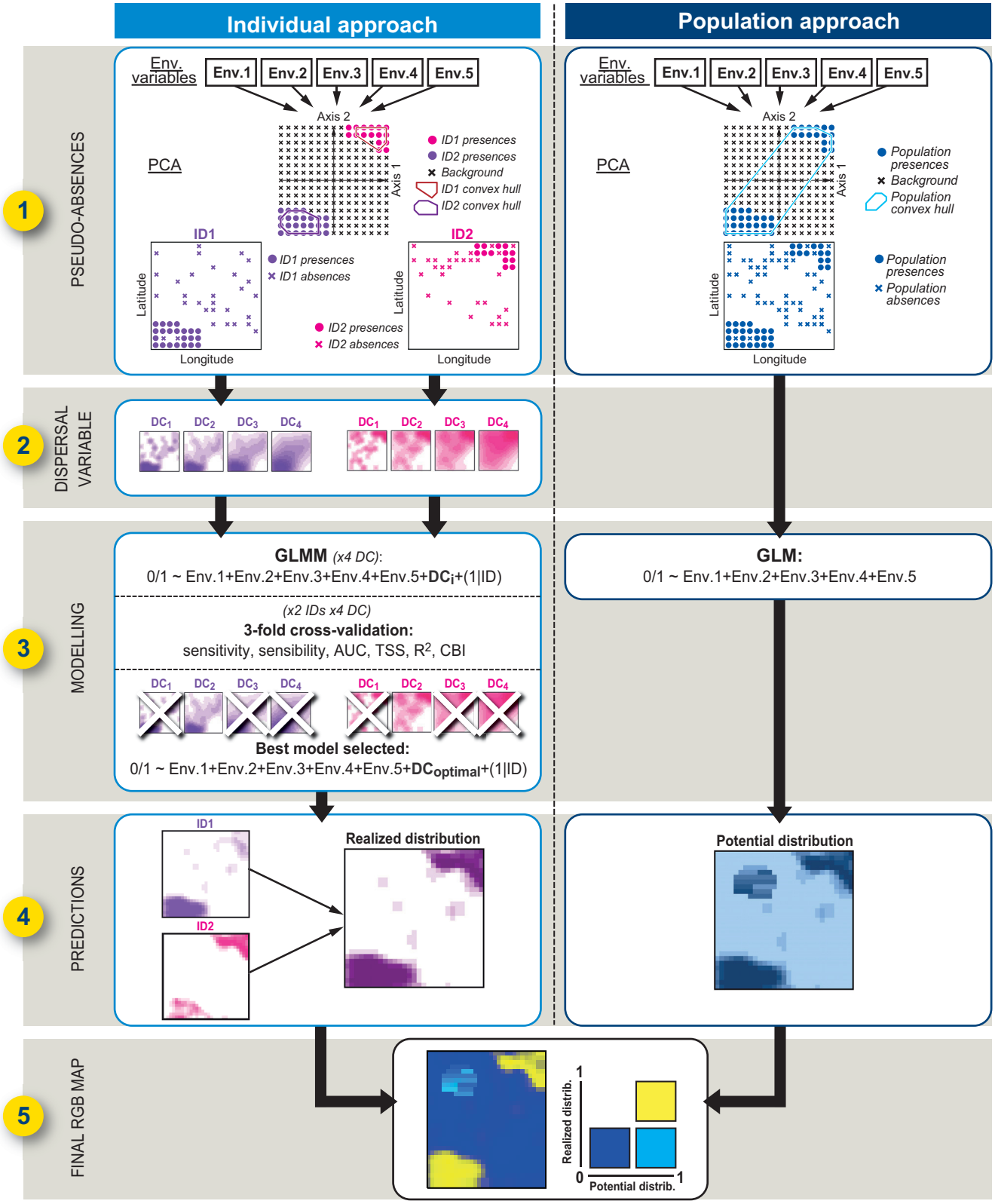

Figure 2. Conceptual diagram of the methodology used in this study, including the five steps described in the Methods section. The diagram was simplified to two individuals (ID1 and ID2) and four dispersal coefficients (DC). Env. refers to the five environmental predictors.

section for the *DC* choice). Such procedure was not applied to the Population approach.

Spatial modelling

Population approach via GLMs

A three-fold cross-validation was used by partitioning the dataset of each run into the training (2/3 of the data) and the validation dataset (1/3). Ten binomial GLMs (one model for each pseudo-absence dataset) were performed on the training dataset using the presence of turtles (1: presence versus 0: pseudo-absence) as a response variable, with a logistic link function (Fig. 2, step 3 right). The ten GLMs included five predictors: slope, slope variance, concavity, distance to rocks and bathymetry that were scaled between 0 and 1, and collinearity was checked using the variance inflation factor (below four). Model evaluation was done on the validation dataset using the five performance metrics calculated for each model (10 GLMs): the continuous Boyce index (CBI, Boyce et al. 2002, Hirzel et al. 2006), the area under the curve (AUC), the sensitivity, the specificity and the true skill statistics (TSS). The models goodness-of-fit was also assessed using the adjusted R^2 (marginal and conditional for the mixed models).

Spatial autocorrelation of the regression residuals was tested using a variogram for each model as well as Moran's I test. To test the influence of different tracking durations and individual sample sizes on the model outputs, the dataset was reduced to 60 days of tracking and 120 locations per individual. The reduced dataset was then subsampled to two daily locations per individual. The performance metrics, response curves and predictions maps were then compared between the full and the reduced datasets to assess for spatial autocorrelation and test the influence of a subsampling procedure (Supporting information).

Individual approach via GLMMs

A similar procedure was done for the Individual approach, except that 1) the turtle's ID was included as a random factor into the GLMMs, 2) fidelity-related variable (*DC*), specific to each individual was also incorporated in the model in addition to environmental variables (Fig. 2, step 3 left). The optimal value of the parameter *DC* in the negative exponential dispersal kernel was selected by testing all possible values of *DC* as a predictor (ranging between 0 and 1) into the GLMM for each run, and by selecting the value that optimizes the model's predictive accuracy and which was visually realistic in regards to individuals' occurrences.

The CBI was calculated individually in order to account for inter-individual variability in models' performance. The choice of the optimal *DC* ($DC_{optimal}$) was based on the CBI criterion since it is considered as the most appropriate metric in the case of presence-only observations. A common $DC_{optimal}$ was chosen for all individuals.

Prediction maps

The selected models were then used to generate the predictions at a 5 × 5 m resolution: 30 predictions maps (3 folds for each of the 10 runs) for the Population approach (potential distribution) and 210 individual prediction maps (10 runs × 21 turtles) for the realized distribution (Fig. 2, step 4). The average prediction maps were then generated for both approaches.

All individual predictions (10 runs × 21 turtles) were then stacked into one single map and averaged (the realized distribution). The differences in terms of sample size between the three sectors (n = 10 in sector A versus n = 6 and n = 5 in sectors B and C) together with the inter-individual variability could lead to an underestimation of the individual presence in some sectors when using the averaged realized map alone. The maximum prediction map derived from the individual maps was therefore also calculated. From a biological conservation perspective, the use of the maximum probabilities value ensures detection of areas containing a single individual, what makes sense in the case of rare and extremely threatened species.

The maps of coefficient of variation (CV) were finally generated to assess the uncertainty of the predictions. For the individual approach, the individual average CVs were first calculated for each turtle based on the 10 individual runs to account for the variability of the pseudo-absence generation. The average CV from the 21 individual CVs was then generated to account for the heterogeneity between the individual predictions. Initially developed to assess variance in species abundance distributions, the Shannon–Weaver index (Shannon and Weaver 1949) was also calculated to account for the inter-individual diversity.

RGB map

Using a standard RGB (red, green, blue) colour space, the spatial predictions of the potential (at the population level) versus the realized (at the individual level from maximum probabilities) distributions were generated from the averaged GLM and GLMM predictions, respectively (Fig. 2, step 5). A similar RGB map was generated by combining the realized distribution from the average probabilities and the Shannon index accounting for inter-individual diversity.

Results

Model selection and performance

Ten *DC* values ranging between 0 and 1 were tested for each turtle, but only seven were retained between 0.004 and 0.03, as values below 0.004 were too low and those above 0.03 too high to make the models converge properly. The CBI varied across the *DC* for the Individual models (Supporting information, based on the smaller CBI range across individuals, range: 0.67–0.99), the selected model had a *DC* of 0.03.

Both approaches showed very little sensitivity to pseudo-absences generation over the different runs. The values of the performance metrics were high (mean range: 0.60–0.99) with little variability (SD range: 0.0005–0.063) for both models (Supporting information). Despite its very good values (CBI > 0.6, positive values indicate a model which predictions are

consistent with the distribution of presences in the evaluation dataset), the CBI showed some variability for the Individual model (CBI range: 0.67–0.99). The Individual model had higher performance values for all metrics except for the CBI. Although their good predictive power, the lowest performance value was for the adjusted R^2 of the population model (mean: 0.6).

Despite a drastic reduction, some spatial autocorrelation was still present after the cross-validation subsampling but with a low sill of 0.03 (Supporting information). Except for the CBI and sensitivity, the performance metrics were slightly higher for the reduced dataset of the Global model (Supporting information). For the individual model, all performance metrics were similar between the reduced and the full dataset (Supporting information).

Response curves

The averaged models (from the 10 simulation runs) contained all environmental predictors for both approaches and all variables were highly significant (Fig. 3, Supporting information). The individual approach also included the dispersal coefficient ($DC_{optimal} = 0.03$) and all variables were significant except the *bathymetry* that had no effect on the probability of turtle presence ($p = 0.892$, Fig. 3). The relationship between turtle presence and the *bathymetry* was negative for the population model. The *DC* had a positive relationship with the probability of presence. For both approaches, the probability of turtle presence was positively correlated to the *slope* and the *concavity*, and negatively correlated to the *slope variance* and the *distance to rocks* (Fig 3). Response curves were similar for both the reduced and the full datasets (Supporting information).

Potential and realized distributions

The map of the realized distribution of sampled individuals (based on maximum values) highlighted three main patches of high probabilities located around the three tagging sites (Fig. 4a). The lowest probabilities were located where pseudo-absences were generated, and conversely, higher probabilities fit well the presence locations (Fig. 4b–d).

On contrast, for the potential distribution, suitable habitats were much more evenly distributed along the coast (Fig. 4e). Given the strong inter-individual variability, the CV was lower for the potential compared to the realized distribution (Supporting information). The prediction maps were similar between the reduced and the full dataset (Supporting information).

Combined RGB predictions map

The combination of both realized (from maximum probabilities) and potential predictions produced the RGB map

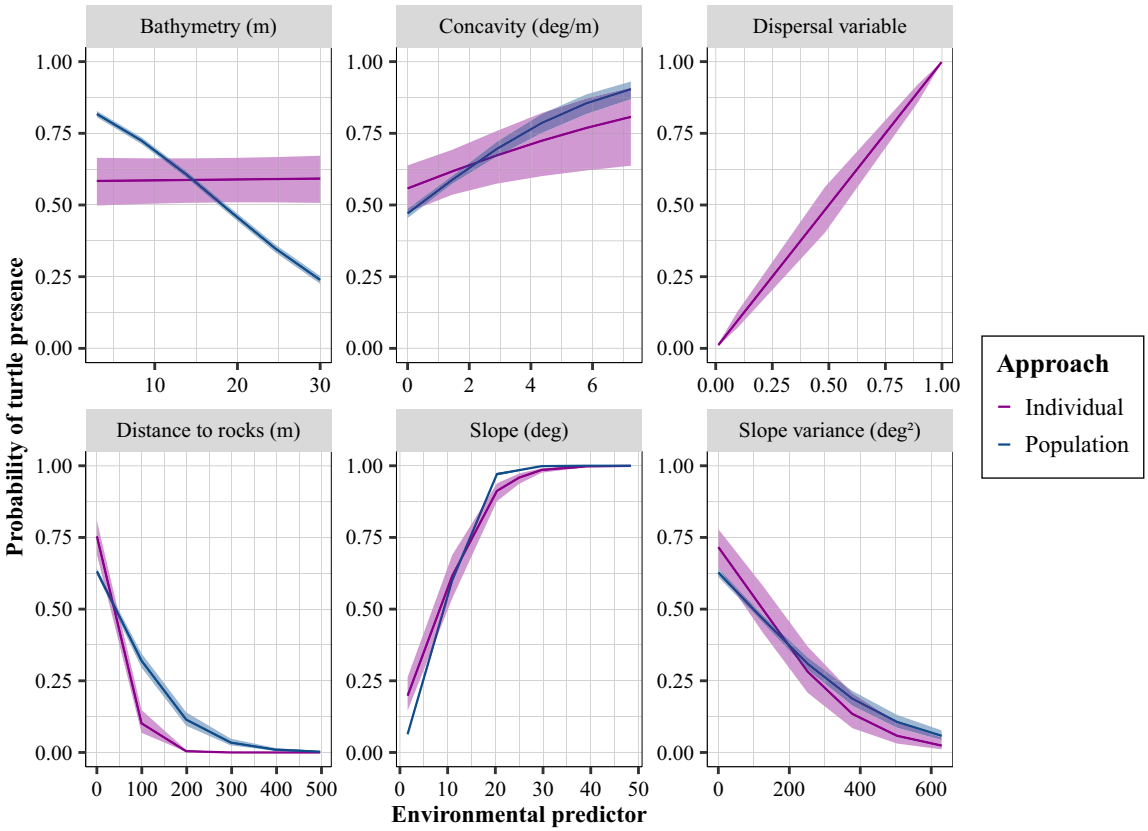

Figure 3. Response curves of each environmental predictor from the linear models for both approaches (individual in purple and population in blue).

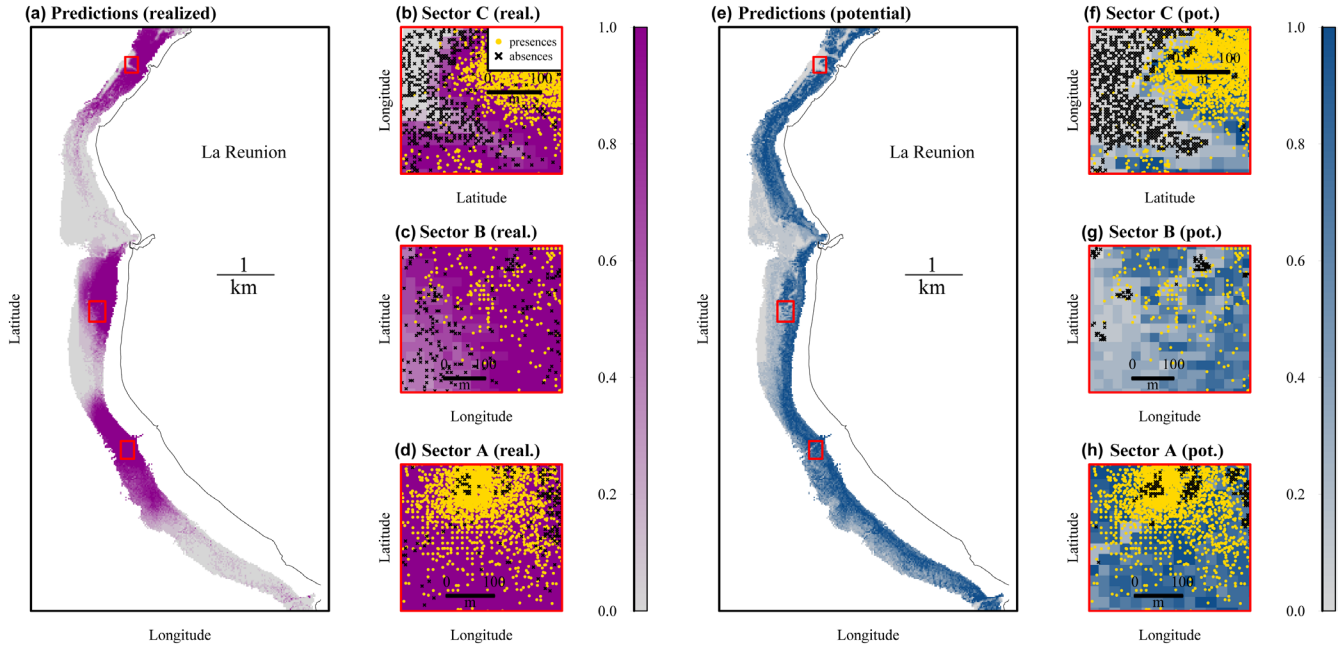

Figure 4. Prediction maps of the realized ('real.', a–d, based on maximum values from the Individual model) and potential ('pot.', e–h) distributions (from the population model) with the associated presences (yellow dots) and pseudo-absences (black crosses). The colour bar ranges from 0 (low probability of turtle presence) to 1 (high probability of turtle presence). Red squares in (a) and (e) indicate the zoom locations (respectively b–d and f–h).

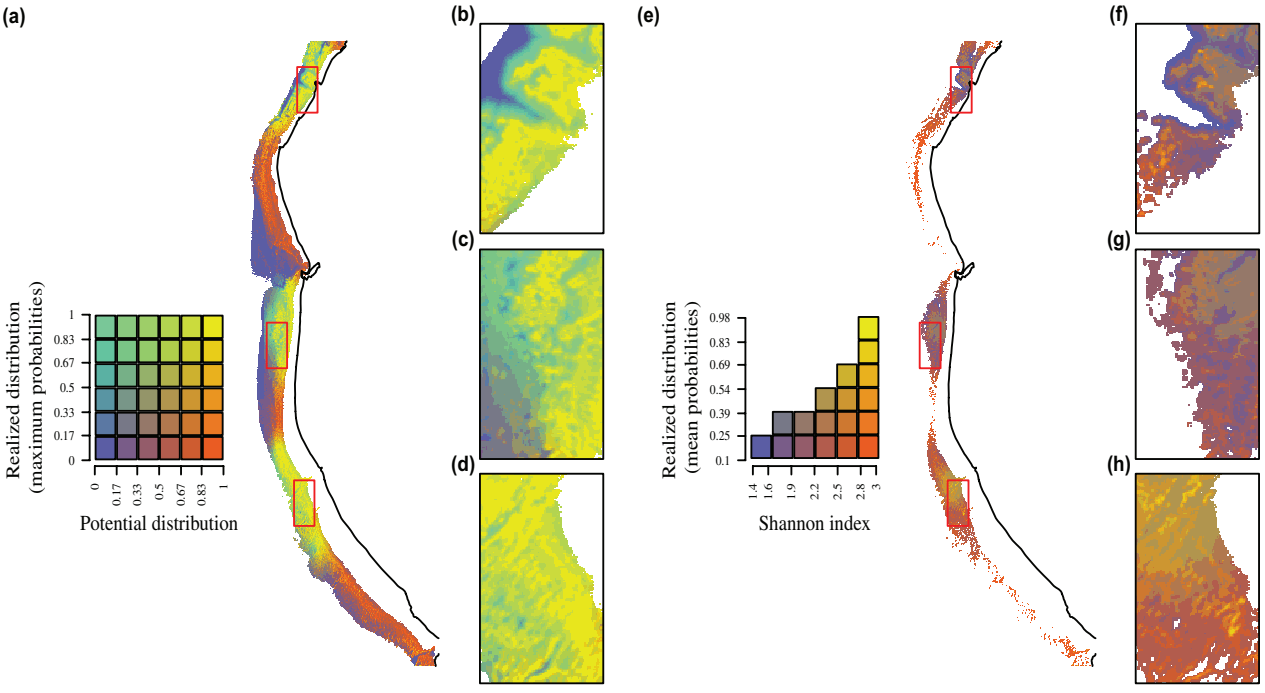

Figure 5. RGB map of the (a–d) realized distribution (based on maximum probabilities) versus potential distribution predicted for the target species. (e–h) RGB map of the average realized distribution versus Shannon index. Yellow colours in the three-dimensional RGB (red, green, blue in a–d) colour space represent areas with a high likelihood to be occupied by the green turtle from both the Individual and Population models. Orange colours represent areas with a high likelihood to be occupied at the population level – or high Shannon index but not at the individual level. Conversely, cyan colours represent areas with a high likelihood to be occupied at the individual level but low at the population level. Similarly, yellow colours in (e–h) refer to high probabilities of average realized distribution and high Shannon index (large number of individuals), and blue colours represent areas with low likelihood to be occupied.

presented in Fig. 5. Areas where both models indicated high probabilities of presence (yellow dots) were mainly located in the three tagging sites. Areas of high realized probabilities and low potential probabilities (cyan areas) occurred at the edges of the three tagging locations, mainly in close proximity to deeper waters (Fig. 5a–d). Conversely, areas of low realized probabilities and high potential probabilities (orange areas) were located in the shallowest waters, southern of sector C and between the three tagging locations. The RGB map based on the averaged realized distribution in relation to the Shannon index indicated a strong inter-individual diversity (orange areas) in Sector A compared to the two other sectors (Fig. 5f–g).

Discussion

Our study provides the first species distribution modelling framework integrating individual-level information provided by tracking data. Predictions from species occurrences for both the potential and realized distributions offer new possibilities to address key questions in biogeography, behavioural and biological conservation.

Biogeographic and ecological applications

Our framework provides a tool to better understand the mechanisms underlying geographic ranges. Our potential distribution based only on environmental predictors showed larger distribution range compared to the individual model. The geographic predictor (the dispersal coefficient) included in the individual model restricts the predictions in the vicinity of occurrences (Meentemeyer et al. 2008). Such approach has been strongly recommended when modelling realized distributions (Lobo et al. 2006, Hattab et al. 2017), and should be systematically implemented to take into account disease dispersal, movement capacities, speed pollination, physical barriers (e.g. lakes, mountains). In our study, the higher values of the performance metrics derived from the optimal dispersal coefficient were used to model the realized distribution of the target species with higher accuracy and reliability, taking into account inter-individual plasticity and spatial fidelity. Despite the swimming capacity of the target species to navigate between the three tagging sites, 89% of the individuals remained in close proximity to their catch and release location. When looking at the individual tracks, this low connectivity between tagging sites (< 10 km apart) reinforced the necessity of using the distance constraint in the case of juvenile green turtles. One interesting contribution would be to see if individuals differ in their responses by incorporating each environmental predictor as a random slope in the GLMM besides the individual on the intercept.

Intrinsic (e.g. individual experience, personality) and biotic factors (e.g. intra-specific competition, predation) can also explain the affinities for some sectors highlighted by the Individual but not the Population approach. For instance, in the present study, the Ermitage site is known to host the highest density of sea turtles (Jean et al. 2010) as it is characterized by the presence of a channel (~50 m width) connecting the outer reef slope to the shallow reef flat habitats (Chambault et al. 2020a). This particular site may provide a more suitable habitat for the development of red algae, the main food resource of green turtles in this region (Ciccione 2001), explaining why we found a higher inter-individual diversity in this area.

In contrast, the population approach can lead to higher estimations in areas that were actually not used by the tracked animals. Although similar environmental conditions can be found outside individuals' realized distribution, some animals might also avoid such 'potentially favourable areas' for biotic reasons (Soberón and Peterson 2005). The comparison between potential and realized distributions using individual tracking data therefore provides new possibilities to investigate complex behavioural processes observed in a wide variety of animals, i.e. intra- and inter-specific competition, predation avoidance, anthropogenic effects.

Conservation and management implications

The use of SDMs within the context of conservation planning has also increased in the past decade (Dawson et al. 2011, Robinson et al. 2011, Cuddington et al. 2013), but the accuracy of the potential distribution of a species derived from SDMs has been challenged, especially when studying rare species (Ochoa-Ochoa et al. 2016). By combining a SDM and a behavioural model at the individual level, the proposed framework provides a more accurate modelling tool to supply managers and stakeholders in their decisions. For species of high conservation interest, our approach provides a map of flexibility by comparing the realized and the potential distributions, accounting for the spatial dynamics of the target species, providing a more accurate suitability map. For example, discrepancies between realized and potential distributions can misdirect management measures. Although the aim of a Nature Reserve is to protect the key species effectively using the minimal possible space (Wilson et al. 2005), the delineation of a protected area based on a potential distribution that covers a broader surface compared to the realized distribution might lead to the implementation of ineffective conservation strategies. By accounting for the individual spatial dynamics, the use of behavioural data derived from satellite tracking will therefore make a significant methodological contribution to the conservation assessment and reserve design of many sensitive taxa. One possible output could be to gradually categorize the potentially suitable areas into different categories based on the combination map of the realized and potential distributions as follows:

- High priority zones located in favourable areas for both the realized and the potential distributions referring to yellow areas in Fig. 5 – such areas having higher priority in 5e than 5a due to a larger number of individuals.
- Medium priority zones for favourable areas for the potential but not for the realized distribution referring to orange

areas in Fig. 5 – such areas having higher priority in Fig. 5e than 5a.
– And low priority zones for areas unfavourable for both the realized and the potential distributions referring to blue areas in Fig. 5.

It is however worth mentioning that our study was limited to one particular life stage of the green turtles which commonly occupies a restricted geographical range before reaching sexual maturity. This species being highly migrant (partly in the open ocean), our framework needs to be used complementary to other approaches in order to protect all life stages of this species which might occupy larger ecological niches.

Limitations and recommendations

Model performance relies on sample size (Papeş and Gaubert 2007, Pearson et al. 2007), and it increases with increasing sample size under constant prevalence (Proosdij et al. 2016). Stockwell and Peterson (2002) showed a drastic decrease in model performance for sample sizes lower than 20 occurrences. In our study, we fixed the minimum number of locations to 50 for each individual, and the mean number of locations recorded per individual was 1047 ± 870 (min–max: 75–4909), being large enough to guarantee reliable predictions that are representative of the tracked animals. Some spatial autocorrelation remained inevitably due to the very fine-scale movements of the individuals (low variogram sill of 0.03, Supporting information). Given the known travel speed of this species (on average 0.15 ± 0.2 km h^{-1}) and its short dive durations at coastal sites (~10–20 min), we assumed that the number of daily locations (mean ± SD: 13 ± 11) was low enough to prevent from a strong spatial autocorrelation between locations (high potential of turtles' occurrences between the received GPS locations) that could violate the hypothesis of independence between observations.

One way to avoid unreliable predictions when combining mixed models to SDMs is to record a sufficient number of locations to be representative of the movement pattern of each animal. The minimum tracking length that should be considered to minimize prediction errors has not been investigated yet in the literature. In our case, the high values of the performance metrics found in our study confirmed the strong predictive power of our models. A complementary analysis based on a reduced dataset using the same tracking duration and an identical sample size for each individual was also investigated in our study, and it showed similar results, confirming the robustness of our approach (Supporting information). In addition to classical performance metrics commonly used in SDMs, the CBI was also used here and enabled to assess how much model predictions differ from random distribution of the observed presences across the prediction gradients (Boyce et al. 2002). The high CBI values derived from our models reinforce their good predictive performance (Hirzel et al. 2006).

The map of the realized distribution was also dependent on tagging locations. Three main tagging sites were used in this study based on prior investigation of the study area (ground truth data) in order to maximize the number of animals captured. One possible limitation is that some individuals tracked from a different site might have led to a different map of the realized distribution, but it is unlikely since the sampling design included a large number of individuals tracked from distinct areas known to be green turtles' hotspots, being representative of the species' range. The lack of random sampling design was therefore compensated by a good representativeness of the population dispersal based on previous studies (Jean et al. 2010, Chassagneux et al. 2013) and periodic at sea observations of the study area, increasing the confidence in our results. However, our framework could not be applied to data collected from a random sampling design which is not representative of the species distribution. Both approaches need to be analysed jointly, especially to help identifying new tagging locations to track additional animals.

The quality and relevance of SDMs depend on the choice of adequate environmental variables, which are in turn constrained by the availability of data sources, often scarce, old, fragmented, unreliable and incompatible for use with high-resolution GPS tracking data. Although turtles' movements are commonly related to dynamic rather than static variables, we restricted the environmental predictors to static variables that matched the fine scale resolution of the study region and that were in agreement with our main hypotheses. Excluding dynamic variables in coastal foraging ground is also not that problematic since oceanographic conditions are quite stable in the tropics and environmental data rarely available at such a fine scale (~1 km width × 11 km length). The use of cutting-edge remote sensing sensors (multi and hyperspectral satellite and aerial images, Lidar data) has therefore shown its interest (high submetric resolutions, exhaustive and extensive spatial coverage), and is a great contribution to ecological studies (Bajjouk et al. 2019).

Conclusion

While traditional SDM approaches commonly neglect the individual information provided by tracking data, the full potential of such individual data is never exploited when modelling species distributions. By running classical SDMs (population approach) with mixed models (individual approach) including individuals as a random factor and a dispersal coefficient accounting for spatial fidelity, we proposed an innovative five-steps framework to predict the potential and realized distributions of mobile species from tracking data. Both approaches showed a strong predictive power but differed geographically in terms of predictions. Our innovative way to combine predictions from both approaches into a single map provides a unique scientific baseline to support conservation planning and management of many taxa. Our framework is easy to implement and brings new opportunities to exploit existing tracking dataset, while addressing key ecological questions such as inter-individual plasticity and social interactions or conservation issues.

Data availability statement

The environmental data is available from Sextant Digital repositiory <https://doi.org/10.12770/7b316584-a52a-49c5-a819-9b78bb9f2159>. Also the location data of the turtles is available from Sextant <https://doi.org/10.12770/d583c380-af34-44cf-9f75-9afee2a6dacc>. (Chambault et al. 2020b).

Acknowledgements – The analyses were carried out within the framework of the NEXT project with the support of IFREMER, Institut Océanographique de Monaco, Fondation Albert I[er] and the Prince of Monaco. The study was also part of the following projects: DYMATURE, HATOCAM and CHARC carried out by IFREMER, Kelonia, CEDTM and IRD. This study was also undertaken within the framework of the 'National Action Plan for Sea Turtles in the French territories of the south-west Indian Ocean'. The authors also thank the logistical support of OceanObs, BNOI/ONCFS and Réserve Naturelle Marine de La Réunion, and to the staff of the Délégation Océan Indien of IFREMER. We are also grateful to Benjamin Robira for its help on the analyses, Pierre Lopez for designing the conceptual scheme and George Hughes for helping with the English.
Funding – The study was financed by IFREMER, DEAL La Réunion, CEDTM, Kelonia, Conseil Régional de La Réunion, IFRECOR, BNOI/ONCFS, Fondation Crédit Agricole, IRD and the French Ministry. PC was funded by IFREMER.
Author contributions – PC and TH wrote the manuscript. PC performed the data analysis. SC, JB, CJ, MD, PM and KB designed the experiment. SC, JB, KB, CJ, MD collected the data. TB and PM: environmental data pre-processing, methodology and conceptualisation. PC, TH, PM, MD, CJ, KB, SC, TB and JB shared the responsibility for contributing to the final version of the manuscript.
Conflict of interests – The authors declare that they have no competing interest.
Ethics approval and consent to participate – This study meets the legal requirements of the French territories where this work was carried out, and follows all institutional guidelines. The protocol was approved by the 'Conseil National de la Protection de la Nature' (CNPN, <www.conservation-nature.com>) and the French Ministry of Ecology acting as an ethics committee in French territories. The committee reference numbers were the following: the projects CHARC (no. 19/12/2012) and HATOCAM (no. 2017-01) in La Reunion.

References

Abrahms, B. et al. 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. – Divers. Distrib. 25: 1182–1193.

Alamgir, M. et al. 2015. Modelling spatial distribution of critically endangered Asian elephant and Hoolock gibbon in Bangladesh forest ecosystems under a changing climate. – Appl. Geogr. 60: 10–19.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecol. Model. 157: 101–118.

Bajjouk, T. et al. 2019. Detection of changes in shallow coral reefs status: towards a spatial approach using hyperspectral and multispectral data. – Ecol. Indic. 96: 174–191.

Balazs, G. H. 1999. Factors to consider in the tagging of sea turtles. – In: Eckert, K. L. et al. (eds), Research and management techniques for the conservation of sea turtles. IUCN/SSC Marine Turtle Specialist Group Publication no.4, pp. 101–109.

Ben Rais Lasram, F. et al. 2020. An open-source framework to model present and future marine species distributions at local scale. – Ecol. Inform. 59: 101130.

Bolnick, D. I. et al. 2002. Measuring individual-level resource specialization. – Ecology 83: 2936–2941.

Bolnick, D. I. et al. 2003. The ecology of individuals: incidence and implications of individual specialization. – Am. Nat. 161: 1–28.

Boyce, M. S. et al. 2002. Evaluating resource selection functions. – Ecol. Model. 157: 281–300.

Brodie, S. et al. 2018. Integrating dynamic subsurface habitat metrics into species distribution models. – Front. Mar. Sci. 5: 219.

Cassini, M. H. 2013. Distribution ecology in conservation biology. – In: Cassini, M. H. (ed.), Distribution ecology: from individual habitat use to species biogeographical range. Springer, pp. 145–162.

Chambault, P. et al. 2020a. Contrasted habitats and individual plasticity drive the fine scale movements of juvenile green turtles in coastal ecosystems. – Mov. Ecol. 8: 1.

Chambault, P. et al. 2020b. Data from: A methodological framework to predict the individual and population-level distributions from tracking data. – Sextant Digital repositiory <https://doi.org/10.12770/7b316584-a52a-49c5-a819-9b78bb9f2159>, <https://doi.org/10.12770/d583c380-af34-44cf-9f75-9afee2a6dacc>.

Chardon, N. I. et al. 2020. Incorporating intraspecific variation into species distribution models improves distribution predictions, but cannot predict species traits for a wide-spread plant species. – Ecography 43: 60–74.

Chassagneux, A. et al. 2013. Unraveling behavioral patterns of foraging hawksbill and green turtles using photo-identification. – Mar. Turtle Newslett. 137: 1.

Ciccione, S. 2001. Autopsie de tortues marines *Chelonia mydas*, retrouvées mortes à la Réunion. – Bull. Phaethon 13: 14–15.

Cuddington, K. et al. 2013. Process-based models are required to manage ecological systems in a changing world. – Ecosphere 4: art20.

Dalleau, M. et al. 2014. The spatial ecology of juvenile loggerhead turtles *Caretta caretta* in the Indian Ocean sheds light on the 'lost years' mystery. – Mar. Biol. 161: 1835–1849.

Dawson, T. P. et al. 2011. Beyond predictions: biodiversity conservation in a changing climate. – Science 332: 53–58.

Domisch, S. et al. 2019. Spatially explicit species distribution models: a missed opportunity in conservation planning? – Divers. Distrib. 25: 758–769.

Dujon, A. M. et al. 2014. The accuracy of Fastloc-GPS locations and implications for animal tracking. – Methods Ecol. Evol. 5: 1162–1169.

Dujon, A. M. et al. 2018. Complex movement patterns by foraging loggerhead sea turtles outside the breeding season identified using Argos-linked Fastloc-Global Positioning System. – Mar. Ecol. 39: e12489.

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

Erauskin-Extramiana, M. et al. 2019. Large-scale distribution of tuna species in a warming ocean. – Global Change Biol. 25: 2043–2060.

Gaspar, P. and Lalire, M. 2017. A model for simulating the active dispersal of juvenile sea turtles with a case study on western Pacific leatherback turtles. – PLoS One 12: e0181595.
Gaspar, P. et al. 2020. Lagrangian modeling of the active dispersal of juvenile leatherback turtles in the North Atlantic Ocean. – EGU General Assembly 2020. <https://meetingorganizer.copernicus.org/EGU2020/displays/35978>.
Hällfors, M. H. et al. 2016. Addressing potential local adaptation in species distribution models: implications for conservation under climate change. – Ecol. Appl. 26: 1154–1169.
Hattab, T. and Lenoir, J. 2016. iSDM: invasive species distribution modelling. – R package ver. 1.0. <https://CRAN.R-project.org/package=iSDM>. in press.
Hattab, T. et al. 2017. A unified framework to model the potential and realized distributions of invasive species within the invaded range. – Divers. Distrib. 23: 806–819.
Hays, G. C. et al. 2002. Behavioural plasticity in a large marine herbivore: contrasting patterns of depth utilisation between two green turtle *Chelonia mydas* populations. – Mar. Biol. 141: 985–990.
Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species presences. – Ecol. Model. 199: 142–152.
Hutchinson, G. 1957. – Cold Spring Harbor Symposium on quantitative biology. Concluding remarks. 22: 415–427
Iturbide, M. et al. 2015. A framework for species distribution modelling with improved pseudo-absence generation. – Ecol. Model. 312: 166–174.
Jean, C. et al. 2010. Ultralight aircraft surveys reveal marine turtle population increases along the west coast of Reunion Island. – Oryx 44: 223–229.
Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – Ecol. Lett. 12: 334–350.
Krüger, L. et al. 2017. Identification of candidate pelagic marine protected areas through a seabird seasonal-, multispecific- and extinction risk-based approach. – Anim. Conserv. 20: 409–424.
Lalire, M. and Gaspar, P. 2019. Modeling the active dispersal of juvenile leatherback turtles in the North Atlantic Ocean. – Mov. Ecol. 7: 7.
Lecocq, T. et al. 2019. Integrating intraspecific differentiation in species distribution models: consequences on projections of current and future climatically suitable areas of species. – Divers. Distrib. 25: 1088–1100.
Lobo, J. M. et al. 2006. Environmental and geographical factors affecting the Iberian distribution of flightless Jekelius species (*Coleoptera*: *Geotrupidae*). – Divers. Distrib. 12: 179–188.
Luschi, P. et al. 2017. Large-scale movements in the oceanic environment identify important foraging areas for loggerheads in central Mediterranean Sea. – Mar. Biol. 165: 4.
Marcer, A. et al. 2013. Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. – Biol. Conserv. 166: 221–230.
Marini, M. Â. et al. 2010. Predicting the occurrence of rare Brazilian birds with species distribution models. – J. Ornithol. 151: 857–866.
Meentemeyer, R. K. et al. 2008. Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. – Ecol. Appl. 18: 377–390.
Montgomery, D. 2007. Design and analysis of experiments, 6th edn. – Wiley, pp. 752.
Mouquet, P. et al. 2015a. Images en pseudo couleurs des fonds marins des zones récifales de la côte ouest de l'île de La Réunion. – <https://geo.data.gouv.fr/fr/datasets/14035532f28f47c6c577c8a99427ffe45fa5d1a0>.
Mouquet, P. et al. 2015b. HYSCORES: Cartographies hyperspectrales appliquées aux écosystèmes coralliens et à leur état de santé. Rapport intermédiaire de fin de campagne d'acquisition. – <http://www.ifremer.fr/depot/lareunion/HYPERSPECTRAL/HYSCORES/REUNION-150923/01_Hyscores_Reunion_20150923_20150922.pdf>.
Ochoa-Ochoa, L. M. et al. 2016. Using one vs. many, sensitivity and uncertainty analyses of species distribution models with focus on conservation area networks. – Ecol. Model. 320: 372–382.
Papeş, M. and Gaubert, P. 2007. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. – Divers. Distrib. 13: 890–902.
Pearson, R. G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. – J. Biogeogr. 34: 102–117.
Proosdij, A. S. J. van et al. 2016. Minimum required number of specimen records to develop accurate species distribution models. – Ecography 39: 542–552.
Putman, N. F. et al. 2020. Predicted distributions and abundances of the sea turtle 'lost years' in the western North Atlantic Ocean. – Ecography 43: 506–517.
Robinson, L. M. et al. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. – Global Ecol. Biogeogr. 20: 789–802.
Scales, K. L. et al. 2017. Should I stay or should I go? Modelling year-round habitat suitability and drivers of residency for fin whales in the California current. – Divers. Distrib. 23: 1204–1215.
Schickele, A. et al. 2020. Modelling European small pelagic fish distribution: methodological insights. – Ecol. Model. 416: 108902.
Schofield, G. et al. 2010. Inter-annual variability in the home range of breeding turtles: Implications for current and future conservation management. – Biol. Conserv. 143: 722–730.
Senay, S. D. et al. 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. – PLoS One 8: e71218.
Shannon, C. and Weaver, W. 1949. The mathematical theory of communication. – Univ. of Illinois Press.
Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. – Biodivers. Inf. 2: 1–10.
Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – Ecol. Model. 148: 1–13.
Thomson, J. A. et al. 2017. Implications of location accuracy and data volume for home range estimation and fine-scale movement analysis: comparing Argos and Fastloc-GPS tracking data. – Mar. Biol. 164: 204.
Varo-Cruz, N. et al. 2016. New findings about the spatial and temporal use of the eastern Atlantic Ocean by large juvenile loggerhead turtles. – Divers. Distrib. 22: 481–492.
Wilson, K. A. et al. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. – Biol. Conserv. 122: 99–112.
Wisz, M. S. et al. 2008. Effects of sample size on the performance of species distribution models. – Divers. Distrib. 14: 763–773.