The Use of a Predictive Habitat Model and a Fuzzy Logic Approach for Marine Management and Planning

Tarek Hattab[1,2]*, Frida Ben Rais Lasram[1], Camille Albouy[3], Chérif Sammari[4], Mohamed Salah Romdhane[1], Philippe Cury[2], Fabien Leprieur[5], François Le Loc'h[2]

1 UR 03AGRO1 Ecosystèmes et Ressources Aquatiques, INAT (Institut National Agronomique de Tunisie), Tunis, Tunisia, 2 UMR 212 Ecosystèmes Marins Exploités, IRD (Institut de Recherche pour le Développement), Sète, France, 3 Département de biologie, Chimie et Géographie, UQAR (Université du Québec à Rimouski), Québec, Canada, 4 Laboratoire du milieu marin, INSTM (Institut National des Sciences et Technologies de la Mer), Salammbô, Tunisia, 5 Laboratoire Ecologie des Systèmes Marins Côtiers UMR 5119, UM2 (Université de Montpellier 2), Montpellier, France

Abstract

Bottom trawl survey data are commonly used as a sampling technique to assess the spatial distribution of commercial species. However, this sampling technique does not always correctly detect a species even when it is present, and this can create significant limitations when fitting species distribution models. In this study, we aim to test the relevance of a mixed methodological approach that combines presence-only and presence-absence distribution models. We illustrate this approach using bottom trawl survey data to model the spatial distributions of 27 commercially targeted marine species. We use an environmentally- and geographically-weighted method to simulate pseudo-absence data. The species distributions are modelled using regression kriging, a technique that explicitly incorporates spatial dependence into predictions. Model outputs are then used to identify areas that met the conservation targets for the deployment of artificial anti-trawling reefs. To achieve this, we propose the use of a fuzzy logic framework that accounts for the uncertainty associated with different model predictions. For each species, the predictive accuracy of the model is classified as 'high'. A better result is observed when a large number of occurrences are used to develop the model. The map resulting from the fuzzy overlay shows that three main areas have a high level of agreement with the conservation criteria. These results align with expert opinion, confirming the relevance of the proposed methodology in this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hattab.tarek@gmail.com

Introduction

Understanding species distributions is essential for conservation planning and forecasting [1], particularly in the present context of stock depletion and species extinction. Species distribution maps play an important role in developing spatial management measures such as the identification of Essential Fish Habitats or the establishment Marine Protected Areas (MPAs) [2,3], which in turn contribute to sustainable ecosystem-based marine management [3–5]. However, information on the true distribution of many marine organisms remains limited, particularly for species that are difficult to detect [6]. Modelling species distributions based on data samples is one solution to address this lack of knowledge. For instance, Species Distribution Models (SDMs) relate species' distributions based on data samples with the associated environmental and geographical characteristics of the surveyed locations [7].

Nevertheless, the perception remains that the distributions of marine species are uncertain and dependent on the sampling process used to generate the models. Bottom trawls are commonly used as a sampling technique to assess the spatial distribution of commercial species and to obtain fisheries-independent abundance data [8]. However, there are some limitations in the species

detections associated with this technique due to a range of factors, e.g., catchability, gear efficiency, and gear-specific selectivity [8]. A common problem in recording species' distribution results from a false absence, which occurs when a species is not available for capture despite occupying the site, or when a species occurs at a site but is simply not captured. Such data can severely limit the fit of many SDMs [9] and can decrease the reliability of prediction models (see [10,11]). The issue of false absences is further complicated by locations with favourable environmental conditions but where species are absent due to biotic interactions, dispersal limitations or fishing pressure. This latter case is particularly critical when modelling commercial exploited species [12]. Yet the problem of imperfect detection when modelling marine species distributions has been rarely mentioned in the literature [13].

Confirmed absences are very difficult to obtain, especially from bottom trawl survey data and for mobile species. A much higher level of sampling effort is required to ensure their reliability relative to presence data [14]. To cope with the lack of confirmed absence data, presence-only models, or profile techniques, have often been used [12]. These models differ from the group discrimination approaches that require presence–absence or

abundance data [15]. Some well-known examples of profile-type models are the Ecological Niche Factor Analysis (ENFA; [16]), the Genetic Algorithm for Rule-Set Prediction [17], and the maximum entropy method [18]. Comparisons between the various SDMs reveal that group discrimination approaches tend to perform better than the profile-type, or presence-only models [19–21]. A common problem of the profile-type models is that their predictions are often overly optimistic, i.e., they predict the species occurring at too many locations [22].

In order to use group discrimination approaches, artificial absence data are increasingly being used in situations where no confirmed absence data are available. Typically called pseudo-absences, artificial absences are generated and inserted into the selected model in lieu of confirmed absence data. The method selected to generate the pseudo-absences is particularly important because it can influence the final quality of the model [21,22]. Several approaches have been suggested for the generation of pseudo-absences: (i) a random selection of absence points across the entire available area (e.g., [21,23]; (ii) a random selection with a geographically-weighted exclusion (e.g., [24]; and (iii) a selection of sampled locations (i.e., occurrence locations for other species) at which the target species has not been recorded (e.g., [5]). However, there are several issues with these approaches. The first two may produce false absences, even in environmentally-favourable areas for the species [25,26]. The latter approach is unsuitable for bottom trawl survey data because it is unlikely that an entire target group of species caught by a particular type of fishing gear will share a similar sampling bias.

To deal with presence-only records, Engler et al. [22] suggested an intermediate methodology between presence-only and presence-absence distribution models. This approach proposed the use of a habitat suitability map as a way of selecting weighted pseudo-absence data points. These points are added to the original presence-only data and used to improve the logistic regression procedure. This mixed method could be more suitable for bottom trawl survey data than the other approaches outlined above.

Distribution models often result in a predicted probability surface that is then translated into a presence-absence classification map for use in different conservation applications [27]. For instance, these presence-absence maps are commonly aggregated (the 'predict first, assemble later' strategy from [28]) to identify areas that will experience the greatest changes in species composition due to climate change [29]. To convert a probability surface into a binary map, a number of threshold selection methods have been proposed [27]. Given the variety of approaches available to generate a dataset, the method chosen can have a dramatic effect on a model's accuracy and its predictions [27,30], as well as the subsequent conservation planning decisions and outcomes [31]. To address the uncertainty in the presence-absence classification map, we propose the use of a fuzzy logic approach that can be applied in situations where vagueness and uncertainty exist. The fuzzy logic approach has several advantages in situations where: (i) there are no clear cut definitions; and (ii) results cannot be categorised as either 0 or 1 [32]. Given its current application in a range of scientific disciplines [33], it is logical to extend the use of fuzzy logic to the threshold selection procedure in the context of SDMs.

Predicting the distributions of commercially targeted marine species is particularly urgent in over-exploited and damaged ecosystems, such as the Gulf of Gabes. With a soft bottom, shallow slope and a high diversity of fishes, the Gulf of Gabes is the most important fishing ground in Tunisia [34,35]. This coastal area supports 60% of the national fishing fleet and contributes 42% of the national annual fish and crustacean production [36]. This

intense fishing activity began in the early 1980s, after which fisheries experienced large fluctuations in landings until the late 1990s when catches began to progressively decline [35]. This was mainly due to the decreasing of demersal stocks caused by intense bottom trawling activities. Depletion rates are so alarming that in the near future the Gulf of Gabes will be subject to a habitat conservation management plan that excludes trawling activities.

In this study, we initially assess the reliability of the mixed approach proposed by Hengl et al. [37]. We combine ENFA predictions for generating pseudo-absences and regression-kriging (RK) for modelling the spatial distribution of 27 commercially-targeted species in the Gulf of Gabes based on bottom trawls survey data. We then propose a fuzzy logic framework to transform modelled probabilities of occurrence into binary predictions of species presence and absence. We illustrate this framework by applying it to the task of identifying areas that meet the conservation targets for the deployment of artificial anti-trawling reefs (AARs) in the Gulf of Gabes.

Materials and Methods

Study Area

The Gulf of Gabes is located in the southern Mediterranean Sea and covers the second widest continental shelf area of this semi-enclosed sea. The Gulf of Gabes has high fisheries productivity and it serves as a feeding and reproduction area for numerous populations of fishes and crustaceans [34]. Indeed, this ecosystem supports one of the most extensive biocenosis of seagrass (Posidonia oceanica; [38]), which constitutes a major nursery site for several marine species [39]. Accordingly, the Gulf of Gabes is one of the most productive ecosystems in the Mediterranean Sea and has great economic and ecological importance [38].

Bottom trawling is the predominant fishing activity in this area and the gear type that has the largest impact on the target demersal fishes [34]. The regular incursions of trawlers into areas that are shallower than their regulated depth have led to the extensive degradation of P. oceanica meadows [40]. Due to the lack of monitoring and surveillance activities carried out by the marine police and fishery authorities, illegal fishing still takes place. Consequently, effective management measures are required to prevent illegal fishing activities.

The proposed fisheries management plan for the Gulf of Gabes includes a perimeter of AARs that combine anti-trawling structures with artificial reefs. The major functions of these structures will be to protect the coastal zone marine ecosystems and species from the mechanical impacts of trawling. This measure will be especially important for high diversity communities, such as the P. oceanica seagrass beds and associated fauna of biological interest [41]. In addition, AARs aim to reduce the fishing mortality of commercial species, to protect nursery areas and juvenile fish, to create new fishing grounds and/or improve existing grounds, and to increase natural productivity [42]. The AARs will be deployed in areas that are chosen by considering both the presence of favourable habitats of commercial species and areas of high-density P. oceanica seagrass beds.

Bottom Trawl Survey Data and Model Variables

Species occurrence data were collected from the Tunisian bottom trawl survey database gathered by the National Institute of Marine Sciences and Technologies (INSTM, Tunisia) onboard the R/V Hannibal. The sampling net used (vertical opening trawl, GOV: 42/55) had a 20 mm diamond stretched mesh at the cod end and a 15 m horizontal opening. The trawl was towed at a speed of 2.9 knots for one hour and sampling areas were fixed
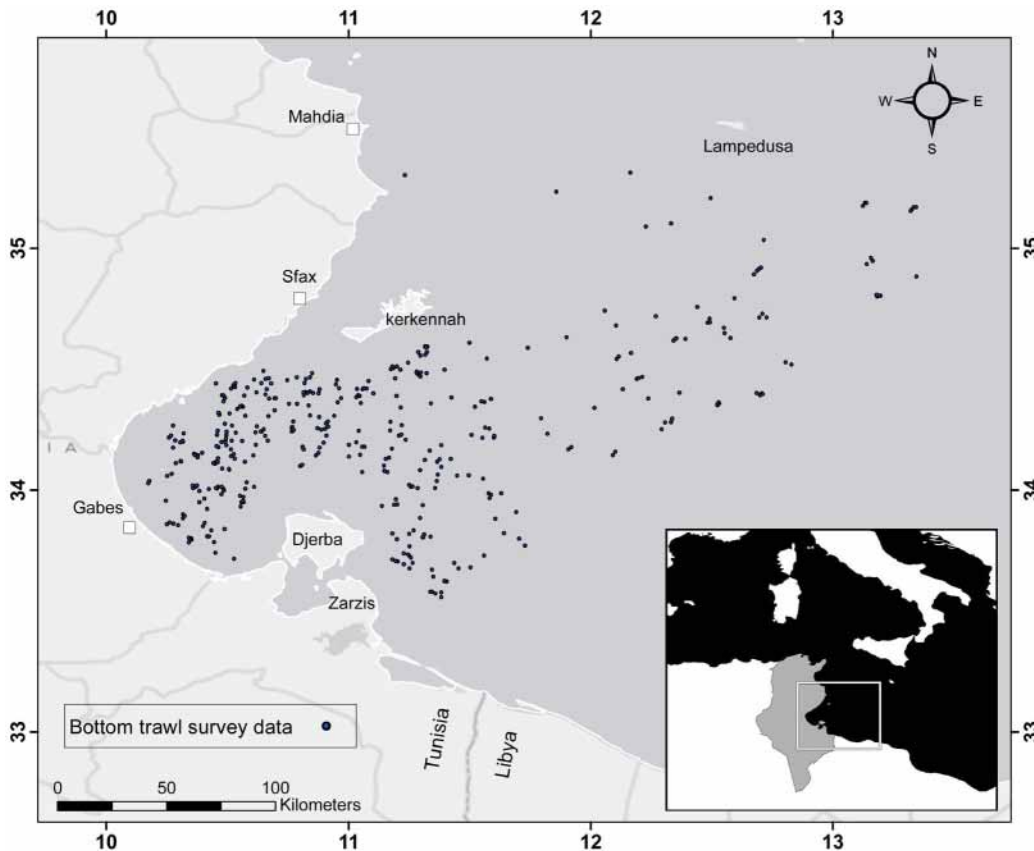
Figure 1. The Gulf of Gabes study area with black points indicating the location of the benthic trawls used in this study.
doi:10.1371/journal.pone.0076430.g001

according to a stratified random sampling based on three depth strata (0–25 m, 25–65 m, and 65–200 m). A total of 360 trawl hauls were completed between 1998 and 2005. Central point geo-referenced position data for each trawl haul and the associated catch contents were extracted from the database (Figure 1).

To build the SDMs, we selected four local habitat variables (depth, slope, aspect, and seafloor type) and one spatial predictor (distance to shore). Previous studies undertaken at a similar spatial extent showed that these variables had a strong influence on species distributions in coastal environments (Figure 2) (e.g., slope [43,44], depth [5,45], aspect [43,44,46], distance to shore [47,48], and seafloor type [5]).

The bathymetry of the Gulf of Gabes was extracted as a digital gridded depth data set from a digital elevation model with a 90 m resolution. The bathymetric elevation data was derived from source soundings collected by the INSTM and referenced to the local tidal datum. The seafloor bathymetric slope and aspect were derived from the bathymetric base map using a 3×3 cell neighbourhood window around the processing cell. Respectively, these represent the rate of change in bathymetry and the azimuthal direction of the steepest slope over the analysis window. The aspect was transformed into two derived variables: Eastness (values close to 1 represent an eastward aspect, while values close to −1 represent a westward aspect) and Northness (values close to 1 represent a northward aspect, while values close to −1 represent a southward aspect).

To develop the map of the distance to shore, Euclidean distances were calculated from the shorelines and islands throughout the study area based on a gridded map with a 90 m

resolution. Release and recapture locations were then sampled using ArcGIS 10 to obtain distance values. Digital seafloor type data were obtained from the INSTM. The original 15 seafloor types were grouped into eight broader seafloor types that represent the relatively distinct physical environments thought to influence the distributions of demersal marine species. We mapped the seafloor type by attributing a seafloor category to the center of each 0.0081 km^2 grid cell.

Selected Species

Since the advent of bottom trawl surveys in the Gulf of Gabes, 152 different species of fishes, cephalopods, and crustaceans have been identified. As small sample sizes pose challenges to any statistical analyses and result in decreased predictive potential, we decided to concentrate on the relatively common species (defined here as species present in >10% of trawl hauls). Having applied this criterion we retained a total of 27 species: 20 fishes, four cephalopods, two decapods, and one stomatopod (Table 1). Collectively, these species constituted 60% of total biomass of the landings and 98% of bottom trawl landings in the Gulf of Gabes [36].

The Modelling Framework

Confirmed absence data were not available for the 27 species selected for this study. As group discrimination approaches usually perform better than presence-only models (as discussed above), we selected a hybrid approach. This approach combines group discrimination and profile techniques to model species distributions (Figure 3) [22,25,37,49].
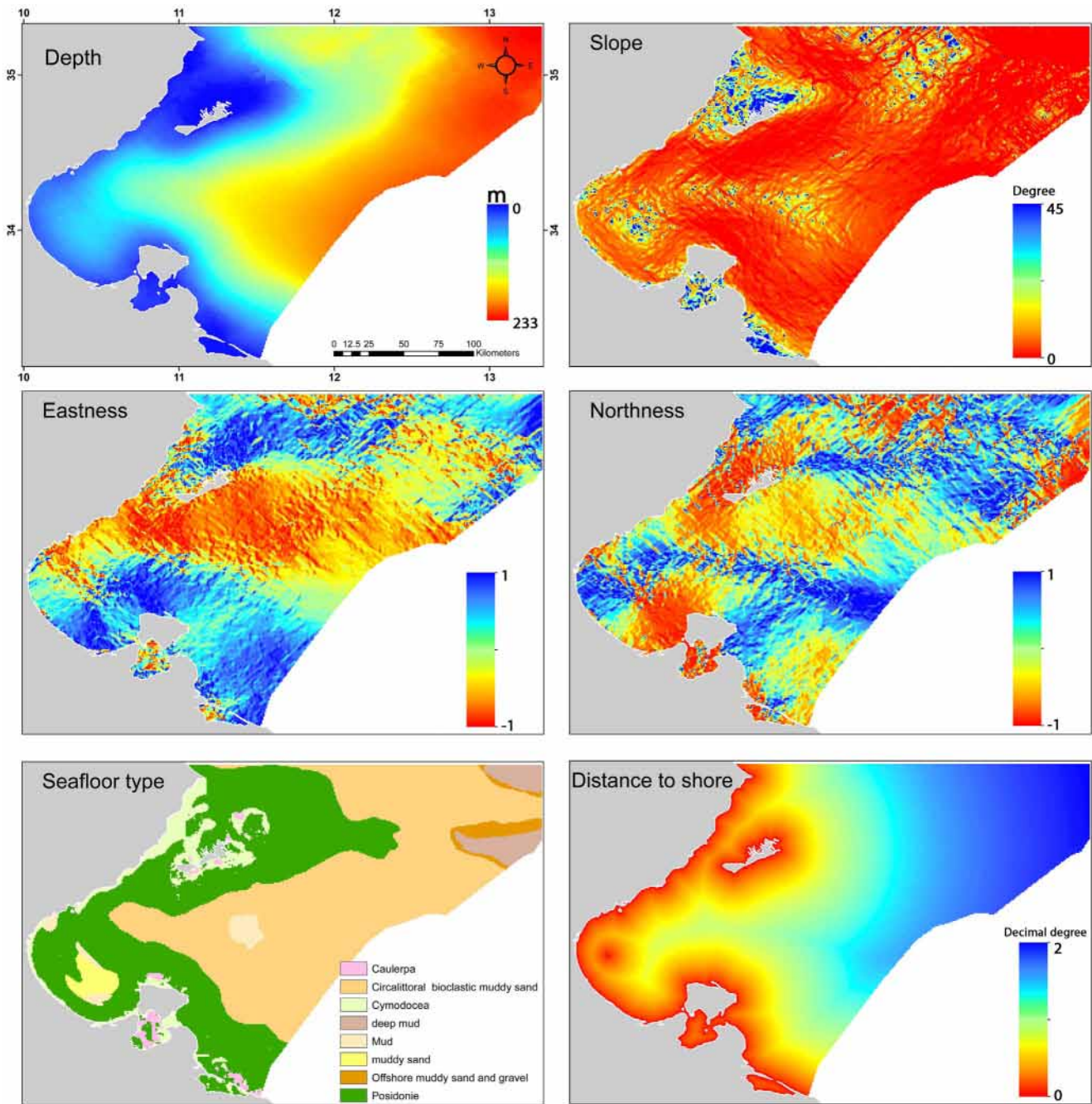
Figure 2. Maps of the six habitat variables (including two derived variables) used to generate the predicted distribution maps.
doi:10.1371/journal.pone.0076430.g002

Presence-only models. ENFA [16] was used to create an habitat suitability map that depicts areas where species are unlikely to occur. ENFA is a specific ordination technique that compares a species' environmental niche and the environmental characteristics of the study area and assigns a degree of suitability to each point on a map (typically from 0 to 100). Thus, it quantifies the dissimilarity between an ecological niche and the ecological space. The first component of the technique, called the marginality factor, is defined as the standardised difference between the centroids of the ecological space and the ecological niche. The second component, the specialisation factors, are successively extracted from the n–1 residual dimensions and represent the

narrowness of the ecological niche relative to the ecological space [16]. ENFA was preferred to another widely used presence-only model, namely maximum entropy method, as several studies recently showed that this technique may lead to spurious inferences (e.g., [50]). Then, ENFA is recognized as one of the best presence-only methods to model habitat suitability for marine species (see [6,51]).

To evaluate the accuracy of the ENFA, we performed a Monte-Carlo randomisation test with 100 permutations. This test for the significance of the marginality factor by randomising the locations of selected species within the study area. At each permutation the ENFA was performed on the random locations and the results

Table 1. Code, scientific name and the number of occurrences of each of the 27 species modelled in this study.

Code	Scientific name	Number of occurrences	Code	Scientific name	Number of occurrences
BALICAR	*Balistes carolinensi*	87	OCTOVUL	*Octopus vulgaris*	77
BOOPBOO	*Boops boops*	299	PAGEERY	*Pagellus erythrinus*	223
DIPLANN	*Diplodus annularis*	190	PAGRCAE	*Pagrus caeruleosticutus*	85
ELEDMOS	*Eledone moschata*	64	PENAKER	*Penaeus kerathurus*	219
ENGRENC	*Engraulis encrasicolus*	47	POMTSAL	*Pomatomus saltator*	51
GOBINIG	*Gobius niger*	61	SCOMSCO	*Scomber scombrus*	95
LITHMOR	*Lithognathus mormyrus*	111	SEPIOFF	*Sepia officinalis*	270
LOLIVUL	*Loligo vulgaris*	134	SERAHEP	*Serranus hepatus*	34
MERLMER	*Merluccius merluccius*	164	SOLEAEG	*Solea aegyptica*	104
METAMON	*Metapenaus monoceros*	46	SPARAUR	*Sparus aurata*	34
MUGICEP	*Mugil cephalus*	38	SPICMAE	*Spicara maena*	62
MULLBAR	*Mullus barbatus*	191	SQUIMAN	*Squilla mantis*	43
MULLSUR	*Mullus surmuletus*	69	TRACTRA	*Trachurus trachurus*	169
MUSTMUS	*Mustelus mustelus*	62			

doi:10.1371/journal.pone.0076430.t001

were then evaluated against the observed locations. We ran the ENFA using the *R* adehabitat package [52,53].

Simulation of pseudo-absences. To generate pseudo-absence data, two methods were used to choose geographic coordinates: (*i*) at random across the Gulf of Gabes; and (*ii*) weighted by ENFA predictions and the geographical location of presence-only records [37]. We undertook both methods with the aim of assessing the level of improvement gained by using the environmentally- and geographically-weighted method as compared with the random method. The weighted method proposed by Hengl *et al.* [37] is based on both the Habitat Suitability Index (HSI, derived through ENFA) and the distance from the observations that are subsequently used to weight pseudo-absence points (Figure 3). Given that HSI values are scaled between 0 and 100, Hengl *et al.* [37] defined the probability distribution (τ) used to generate the pseudo-absence locations as:

$$\tau(x*) = \left[\frac{d_R(x) + (100 - HSI(x))}{2} \right]^2 \qquad (1)$$

where d_R is the normalised distance in the range [0, 100%], i.e., the distance from the observation points divided by the maximum distance. The square term is used to ensure that there are progressively more pseudo-absences at the edge of low HSI and large distances will approximately follow a Poisson distribution. In this way, pseudo-absences are located both in areas of low HSI (unsuitable habitat) and further away from the occurrence locations.

Based on Equation 1, the HSI and the map with buffers around the occurrences were combined to create a weighted map. We then performed the random generation of points with a probability density proportional to the values of the weighted map. To account for the variability arising from this weighted selection, 10 groups of pseudo-absences were generated. This allowed us to assess the stability of the final predictions for different simulations. For each set, the number of simulated pseudo-absences was equal to the number of presences. This is supported by the statistical theory of model-based designs, also known as "D-designs" [37]. According to this theory, the optimal design to minimise

prediction variance is when an equal number of observations are at opposite value extremes [37,54] and there is a higher spreading in the feature space.

A total of 11 groups of pseudo-absences were obtained. One group was generated entirely at random and the remaining 10 groups were weighted by ENFA predictions and the geographical locations of the presence-only records (Figure 3).

Regression modelling. Once the pseudo-absences were simulated, they could then be combined with the occurrence locations to build a regression model to predict the probability distribution of occurrences. Prior to running the regression analysis, the six original habitat predictors were converted to principal components (to reduce their dimensions and the multicollinearity effect) using the Hill-Smith ordination method [55] that deals with mixed variable types (i.e., quantitative and factor).

We used a generalised linear model (GLM) [56] for the regression analysis, assuming a binomial error in the response variable. Model residuals were then analysed by fitting a variogram to assess their level of spatial dependence (Figure 3). Model residuals exhibited no spatial dependence for six species (i.e., the squid, *Loligo vulgaris*; the stomatopod, *Squilla mantis*; the demersal fishes, *Merlucius merlucius*, *Mugil cephalus*, *Mullus barbatus*, and *Mullus surmuletus*; and the pelagic fish, *Trachurus trachurus*). For the remaining species (n = 21), we used logistic Regression Kriging (RK) models that explicitly incorporate spatial dependence into predictions (Figure 3). This method assumes that the model residuals have a spatial structure resulting from either 'model' factors such as incorrectly specified or inadequate predictor variables, or 'real' factors such as biotic processes that cause spatial patterns [57]. It combines the predictions from a regression model along with the resulting kriged residuals [58]. Specifically, the regression modelling was supplemented with the use of variograms to assess the level of spatial dependence among residuals (Figure 3). Regression residuals were then interpolated and added back to the regression estimate (see [58] for more details). Finally, to select the most parsimonious model for each of the selected species, we applied an automatic stepwise model selection using the Akaike Information Criterion [59].
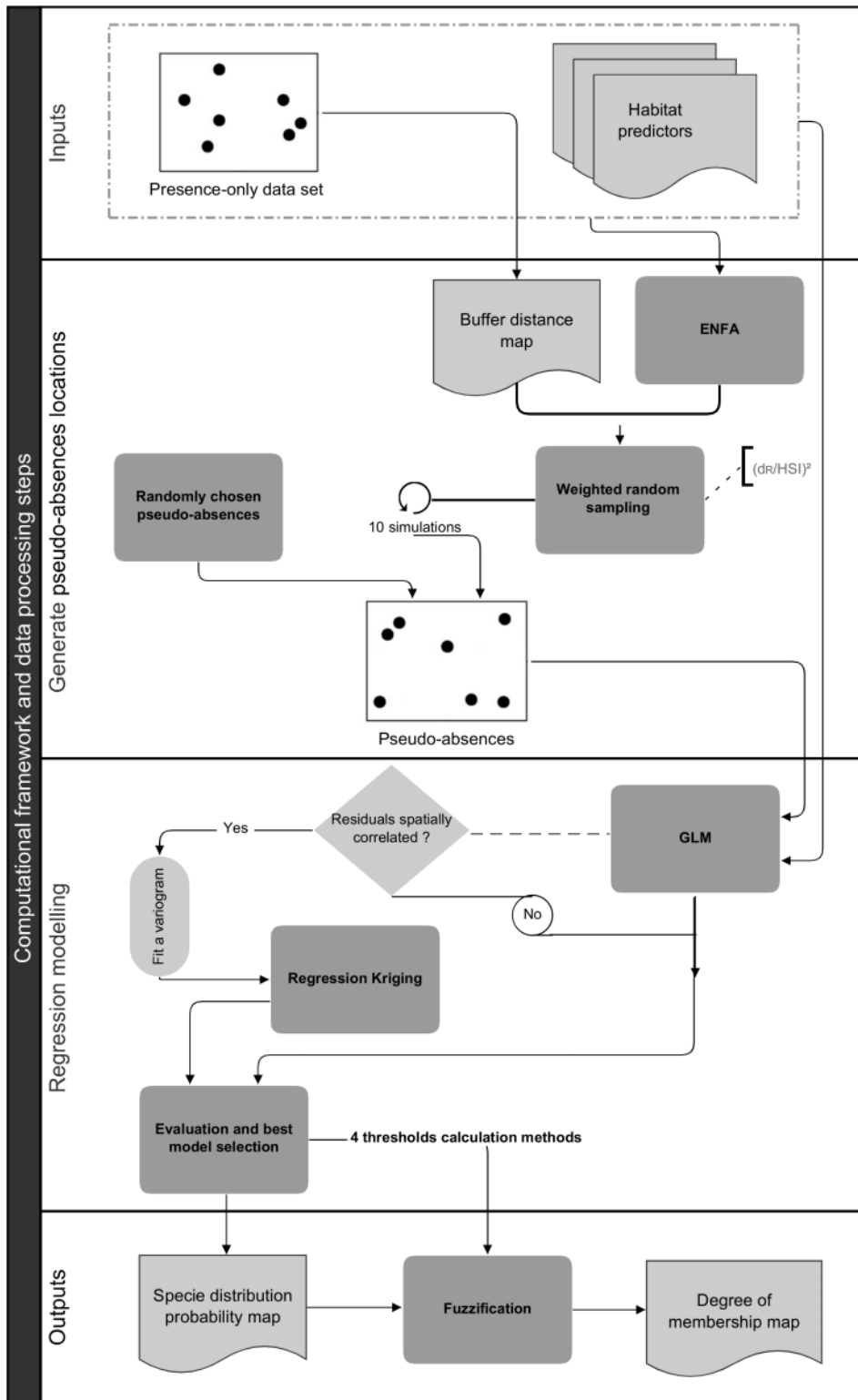
Figure 3. Computational framework and data processing steps.
doi:10.1371/journal.pone.0076430.g003

Model accuracy. For every species, the predictive accuracy of the model was evaluated by a 10-fold cross validation [56]. The receiver operating characteristic curve method was then applied to derive the area under the curve (AUC) index [60] to measure the model's performance.

The simulation of pseudo-absences may generate absences far from the environmental conditions of presences, which may

artificially increases the rate of well-predicted absences and hence the AUC scores [24,25,61]. In addition, the AUC test statistic may not always reflect a model's ability to prioritise areas in terms of their habitat suitability relative to alternative models (e.g., [61,62]). Model assessment was therefore supplemented with the Point Biserial Coefficient (PBC) [20,24,63], the sensitivity (presences correctly predicted as presences), and the specificity metrics (absences correctly predicted as absences) [60]. The PBC was calculated as a Pearson's correlation coefficient between the observation in the occurrence dataset (presence (1) or pseudoabsence (0)) and the prediction and therefore takes into account how far the prediction varies from the observation. An independent examination of the percentage of presence and absence errors was recommended by Lobo *et al* [61] to help in the model selection process according to the researcher's goals, rather than the use of a synthetic measure such as the AUC. Predictions were further inspected visually and compared to plotted occurrence data in order to assess their plausibility.

Finally, we used the Pearson's correlation coefficient to calculate the pairwise correlation between the final predictions maps derived from the 10 groups of pseudo-absences. This allowed us to assess the stability of predictions for each of the different weighted simulations of pseudo-absences.

Each data processing step was completed in *R*, drawing on code developed by Hengl *et al.* [37], and automating the calculation for several species simultaneously. For each modelled species, the regression models, 10-fold cross validation, and evaluation procedures were carried out for the 11 simulated groups of pseudo-absences.

Conservation Planning Procedure

Probability of occurrence maps can be generated using the most accurate model selected for each species from the 11 pseudo-absences groups. Once these maps have been converted to presence-absence data, they can be used to identify areas that meet the conservation targets for selection of AAR deployment sites. These targets specify the inclusion of both favourable habitats for commercially-targeted species and areas of high density *P. oceanica* seagrass beds.

Threshold approaches and fuzzy modelling. The most common method used to convert probabilities of occurrence to presence-absence data is the use of an optimum probability threshold [64]. Different methods have been proposed to select a probability threshold [27]. Among these, the most widely used threshold optimisation criteria are:

(1) *Sens = Spec* - The threshold where sensitivity equals specificity (i.e., where positive observations are equally as likely to be wrong as negative observations).

(2) *Max (Sens+Spec)* - The threshold that maximises the sum of sensitivity and specificity (i.e., it minimises the mean error rates for both positive and negative observations). This threshold is equivalent to finding the point on the receiver operating characteristic curve whose tangent has a slope of one.

(3) *MaxKappa* - The threshold that results in the maximum value of Kappa statistic.

(4) *MaxPCC* - The threshold that results in the maximum percent of correctly classified observations.

For each species, a probability threshold was determined using each of the four optimisation criteria outlined above. As these methods do not provide similar threshold values, the use of one method over another can influence conservation planning

outcomes, e.g., modifying areas that are expected to be suitable for a given species [27]. To avoid the subjective selection of a particular threshold, predicted distributions of selected species were defined by fuzzy sets theory [65]. Fuzzy logic is useful in circumstances that involve uncertainty, imprecision, and vagueness by replacing the sharp boundary between the suitable and non-suitable classes with the concept of a degree of truth (membership). In the fuzzification process, crisp attribute values (probability of presence) are transformed linearly into a common suitability scale (0 to 1) using the fuzzy linear membership function. A membership value of 0 is assigned to the lowest probability threshold value (as calculated using the four criteria outlined above) and a value of 1 to the highest threshold value (see example in Figure 4). Each map cell was assigned a fuzzy membership value resulting from the fuzzy linear membership function.

Fuzzy overlay of conservation criteria. For each species, the favourability value of an area is defined as the degree of membership of that area to the fuzzy set of favourable areas for the species. Fuzzified inputs can be combined together to identify the most favourable area for the majority of commercial species by using a fuzzy operator. The fuzzy AND operator was applied to return the minimum of the fuzzy memberships from the fuzzy input maps. The result of this aggregation is a final fuzzy set expressing the site suitability for all key species (see example of the method in Figure 4). To identify areas with a high density of seagrass beds, a recent map showing seagrass recovery rates was obtained from the INSTM. This map was fuzzified using a fuzzy linear membership function that assigned a membership value of 0 for recovery rates less than 30% and a value of 1 for recovery rates greater than 60%. Finally, the maps that express the fuzzy memberships of site suitability for all key species and seagrass recovery rates were combined together using the fuzzy AND operator. For example, "IF the favourability value of an area for the species 1 IS high, AND the favourability value of an area for the species 2 IS high, etc…, AND the recovery rate of seagrass IS high THEN the area has high conservation criteria".

Results and Discussion

The Monte-Carlo randomisation tests show that the ENFA marginality factor is highly significant for each modelled species (all $p<0.001$). This implies that the habitat occupied by the species modelled differ unequivocally from the average environmental conditions found in the broader study area. This indicates that a species-specific habitat selection process takes place. The ENFA also show that the potential distributions of species are much larger than their realised distributions, based on sampling locations (see examples of three species in Figure 5). In contrast, incorporating pseudo-absence data into logistic RK models results in predicted distributions that are closer to the realised distributions (Figure 5).

When the pseudo-absence data are randomly selected, the potential distribution of the species are less extensive in comparison with those obtained using a weighted selection. This difference can be explained by the fact that a random simulation can select absences between observed occurrences points, therefore generating pseudo-absences in favourable areas. This can subsequently lead to an underestimation in the realised distributions [25,66].

For each method generating pseudo-absences, the distribution of the AUC, PBC, the sensitivity and the specificity metrics is shown in Figure 6. For each species, model accuracy differs according to the method used to generate pseudo-absences. All the models based on the environmentally- and geographically-
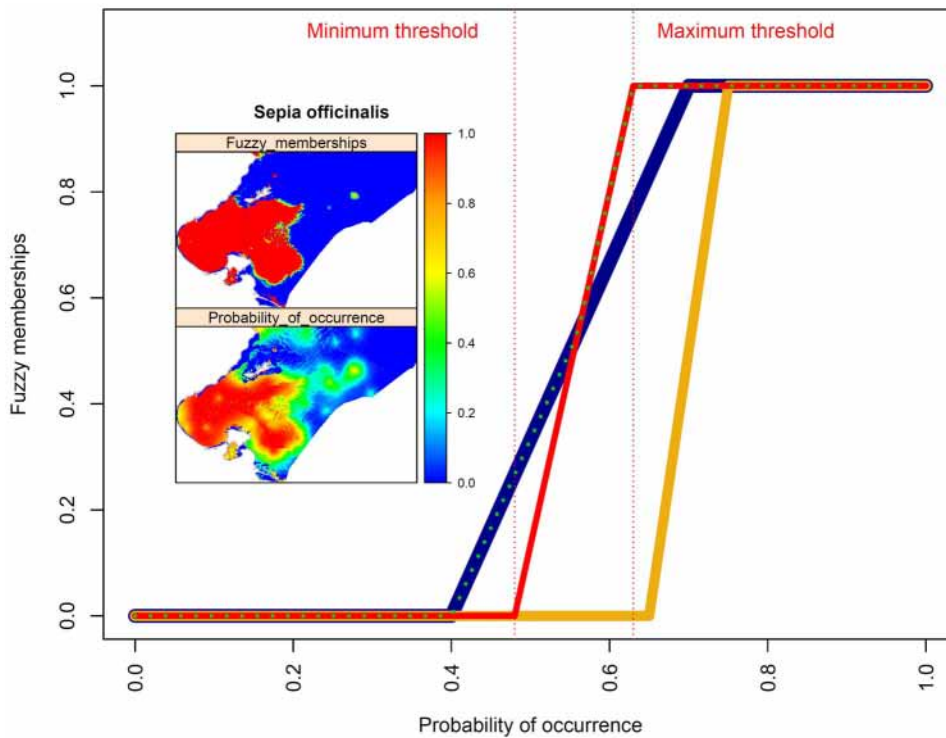
Figure 4. Threshold approaches and fuzzy modelling. An illustration of the fuzzification process performed using a fuzzy linear membership function. The red, orange, and blue curves represent the fuzzy membership sets of three different species. The dotted green line represents the fuzzy AND overlay outcomes. The inset maps display an example of the fuzzification process for the cuttlefish *(Sepia officinalis)* which corresponds with the red curve.

doi:10.1371/journal.pone.0076430.g004

weighted method achieved a high level of accuracy as indicated by the four measures of model accuracy (Mean±standard deviation: 0.9±0.07, 0.65±0.23, 0.87±0.08 and 0.85±0.09 respectively for the AUC, PBC, sensitivity and specificity) (Figure 6). These values were significantly higher (Wilcoxon rank test: $p < 0.001$) than those obtained by models based on the random selection method (Mean±standard deviation: 0.78±0.1, 0.36±0.26, 0.77±0.11 and 0.69±0.1 for the AUC, PBC, sensitivity and specificity, respectively).

Since all the resulting SDMs are based on pseudo-absences, both specificity and AUC scores estimate the degree of accuracy of the absence information used in the model training process. Thus, a high specificity score only implies that most of the data considered as absence data are correctly predicted and does not imply a high performance in the prediction of the unknown true absences. However, the sensitivity and PBC values were higher when the weighted method of generating pseudo-absence was used, implying a high performance of this method in the predictions of the known true presences as compared with the random method.

Models based on the weighted method of generating pseudo-absence data provide significantly better results on average, which aligns with the results of Engler *et al.*, Chefaoui & Lobo and Hengl *et al.* [22,25,37]. It contrasts, however, with the findings of Wisz & Guisan and Barbet-Massin *et al.* [19,26]. Using virtual species, these authors found that randomly-selected pseudo-absences yielded the most reliable species distribution models. However, this may be explained by the fact that both studies used a large number of pseudo-absences (e.g., 10.000 data points), whereas this study used the same number of simulated pseudo-absences as the number of occurrences. Currently, there is no consensus on the

number of pseudo-absences that are required to optimise model predictions. Some authors suggest a ratio of 10:1 (e.g., [25,49]) while others recommend using large numbers of pseudo-absences when they are randomly selected (e.g., [20,26]). Intuitively, it makes sense to generate an equal number of pseudo-absence data points as occurrence data points [58,67] to avoid the bias caused by a presence-absence ratio that is too low [68,69]. Indeed this was the outcome in McPherson *et al.* [70]who found SDMs had the best predictive accuracy when prevalence values (the proportion of data points representing a species' presence) were around 0.5. Therefore, several authors recommend resampling the training data to balance presence and absence data points [70,71].

For each species, the range of pairwise correlations between the final habitat prediction maps (derived from each of the 10 groups of pseudo-absences) varied according to the number of occurrences used to develop the models (Figure 7). The lowest correlation occurs when the number of occurrences is less than 61. The array of evaluation measures, based on all the replicated runs, does not show a clear trend in relation to the number of occurrences used to fit the models. Even with a small number of occurrences, values of AUC and PBC indicate excellent predictive accuracy of the models. However, large fluctuations in the predictive accuracy and the models' subsequent predictions are observed when the models are based on a low number of occurrences. Indeed, the selection of pseudo-absences can induce variability in model predictions when several runs are made with a small set of occurrence data, each run having its own dataset for calibration.

Williams [72] found that the predictive ability of some ecological modelling approaches varies with a species' detectability. While presence-absence approaches generally have higher
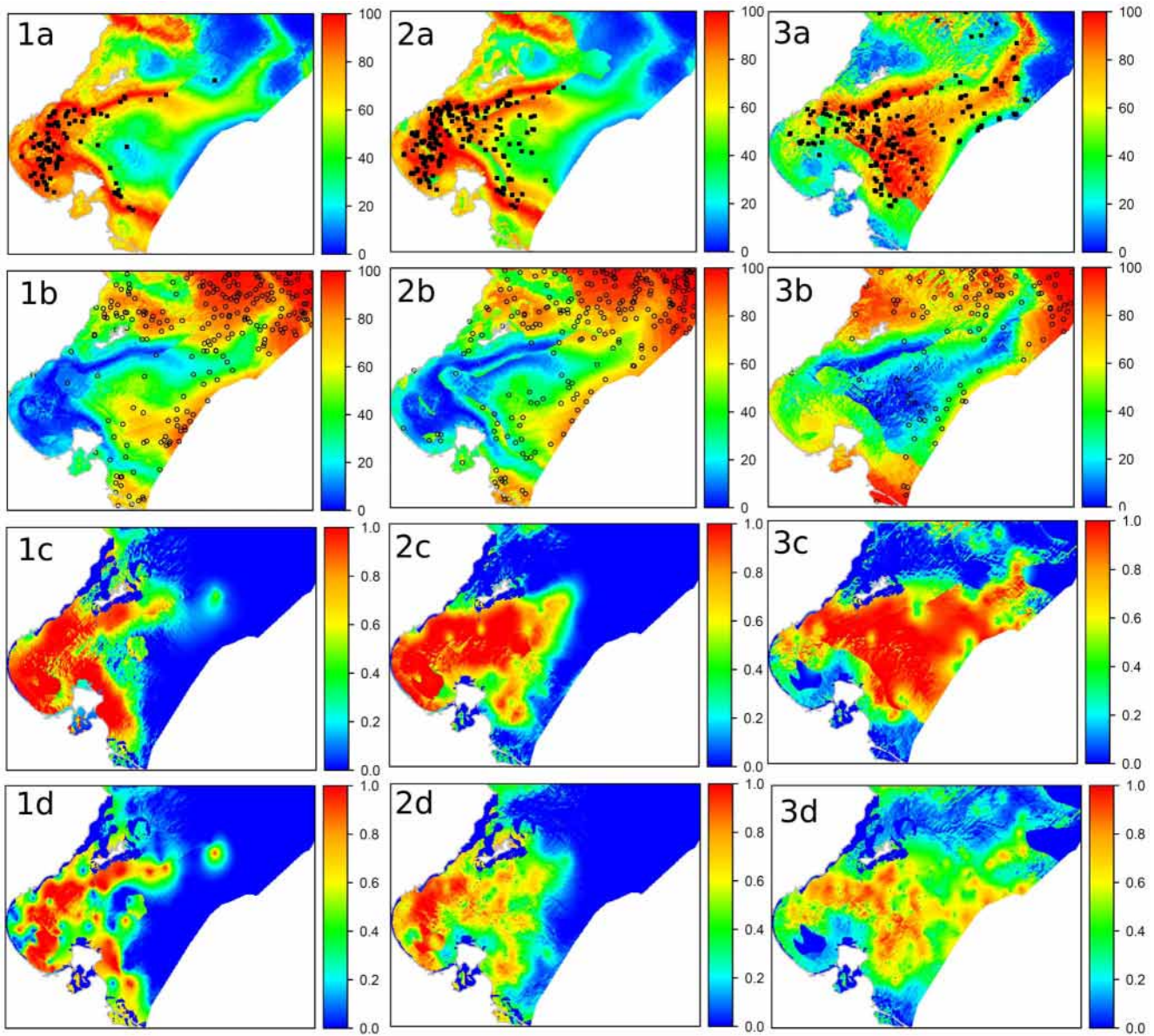
Figure 5. Spatial prediction maps in the Gulf of Gabes for three example species. (1) *Lithognathus mormyrus,* (2) *Penaeus kerathurus,* (3) *Pagellus erythrinus* (a) the habitat suitability index map with presence-only data; (b) the weighted map and the randomly-generated pseudo-absences using the Equation 1; (c) probabilities predicted using the binomial regression-kriging (RK) with a weighted selection of pseudo-absences; and (d) probabilities predicted using the binomial RK with a random selection of pseudo-absences.
doi:10.1371/journal.pone.0076430.g005

predictive abilities for species with a high detectability, they do not perform as well as presence-only approaches when detectability is low. However, for some species such as the pelagic fish species, *Trachurus trachurus*, the fitted model is found to be moderately accurate though it is based on a relatively high number of data points (338 presences/pseudo-absences). This result may be related to the fact that the selected predictors (e.g., depth, seafloor type) have greater ecological significance to model the distribution of demersal and benthic species. [66]. This implies that the performance of SDMs strongly depends on the environmental predictors, which in turn depend of the type of organism that is studied [62]. In addition, the realism and the robustness of models may have been influenced by our automatic variable selection procedure [73]. Indeed, ecologically important variables may have been excluded from the stepwise models and/or non-meaningful

variables may have been incorporated into models. This was particularly the case for some of the studied pelagic species (*Trachurus trachurus*, *Spicara maena*, *Scomber scombrus*). These species were therefore excluded from the conservation planning procedure.

Overall, our results suggest that simulating pseudo-absences with an environmentally and geographically weighted method rather than a purely heuristic approach enhances the accuracy of predictions. This method provides a robust result when a relatively large number of occurrence data points with good spatial coverage are used. The RK method also shows great potential as an approach to incorporate spatial dependence in SDMs, by combining information on species-habitat relationships (i.e., through the deterministic model) and error components. We believe that the combined ENFA and RK method has several
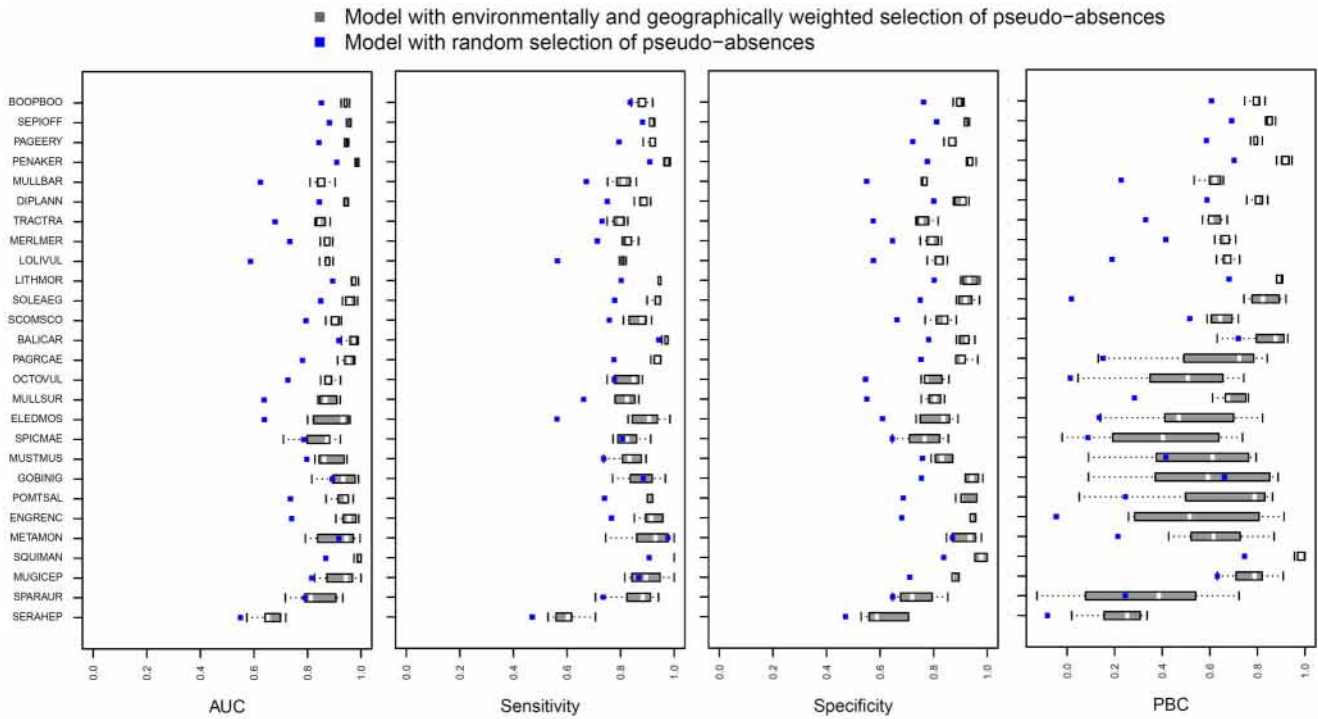
Figure 6. Distributions of the area under the curve (AUC), PBC, sensitivity and specificity values. Calculated AUC, PBC, sensitivity and specificity values for the 27 modelled species (species codes are listed in Table 1). Species are sorted (top to bottom) by decreasing number of occurrences (used to develop the models).
doi:10.1371/journal.pone.0076430.g006

advantages when applied to trawl survey data, especially when addressing their imperfect ability to detect a species. Furthermore, this method applies both the spatial auto-correlation structure and the trend component of the spatial variation to make spatial predictions of species' distributions. This method can be applied to
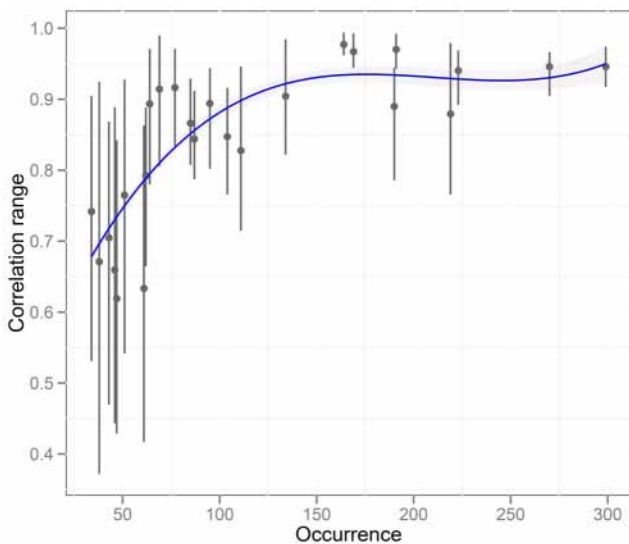

Figure 7. Stability of predictions. The range of Pearson's pairwise correlation coefficients between the final prediction maps according to the number of occurrences used to develop the models (red line: smooth curve fitted by Loess function).
doi:10.1371/journal.pone.0076430.g007

other areas where survey data are available, such as the Medits (International bottom trawl surveys in the Mediterranean) dataset.

The second purpose of this study is to identify the areas required to meet the conservation targets of AARs based on probabilistic predicted maps. In this study we propose a novel method to develop these maps that uses fuzzy sets to address the uncertainty associated with the selection of probability threshold optimisation criteria. Fuzzy logic is interested in capturing partial truths, that is, how to reason about things that are not wholly true or false; while probability is concerned with making predictions about events based on a partial state of knowledge [65]. The fuzzy sets theory is used to transform the probability of presence into a membership degree using not only a single threshold value but several values obtained with different cut-off threshold optimization criteria.

Of the 27 species modelled in this study, an excellent or high predictive accuracy is only found for 12 benthic and demersal species. It was these species that were selected for the fuzzy overlay of conservation criteria. Of these, the 2004–2005 stock assessment results for the Gulf of Gabes report that *Solea aegyptiaca*, *Octopus vulgaris*, and *Sepia officinalis* were fully exploited (Othman Jarboui, personal communication: INSTM) and *Mullus barbatus*, *Pagrus caeruleostictis*, and *Pagellus erythrinus* were overexploited.

The resultant fuzzy overlay map (Figure 8) highlights three main areas that meet the conservation criterion to a high level. The largest area is located south of Kerkennah Island while the remaining two areas are in the coastal region off Mahres Harbour and north of Jerba Island. As well as meeting the AAR conservation criteria, the Kerkennah Island area has already been reported as a biodiversity hotspot for megabenthic fauna [74] and is currently proposed as a potential MPA [75]. The area north of Jerba Island was also proposed as suitable for MPA establishment by Ben Mustapha & Afli [75]. In addition to its dense seagrass
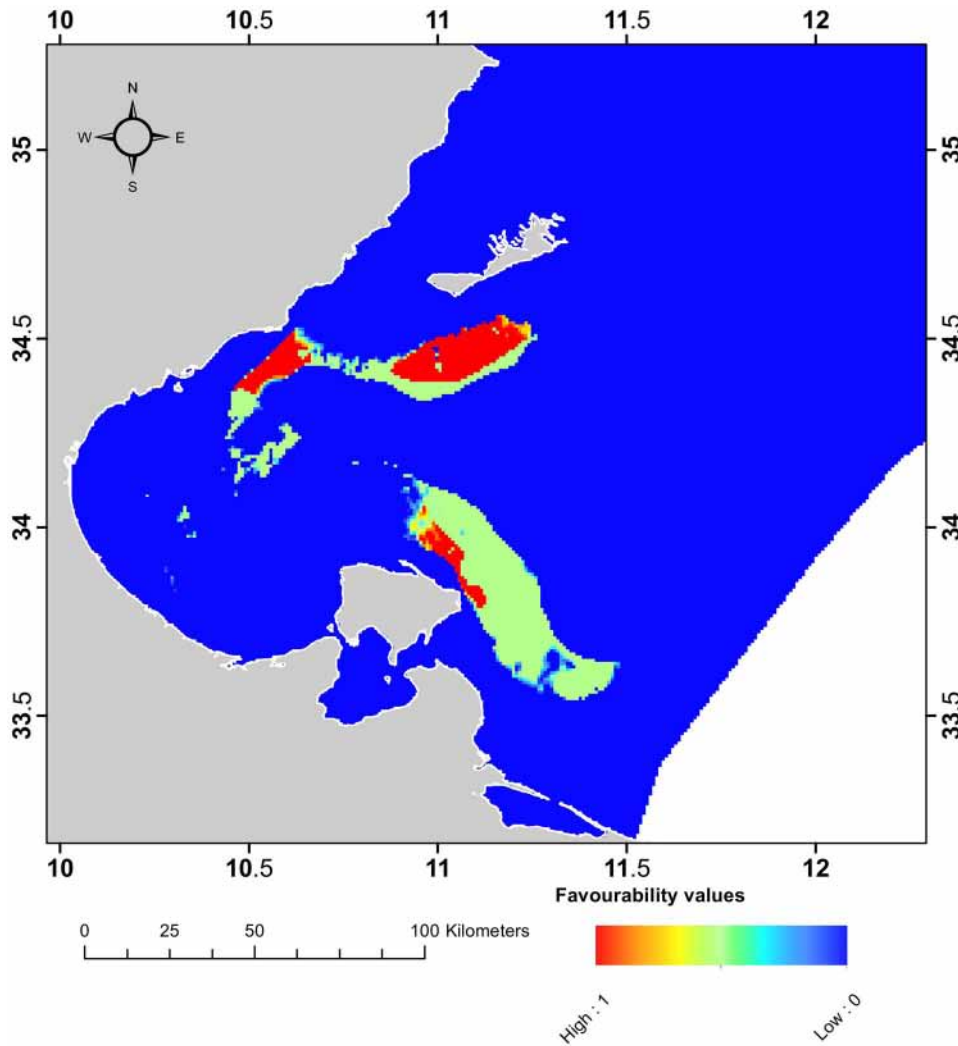

Figure 8. Map of favourability values for the deployment of AARs.
doi:10.1371/journal.pone.0076430.g008

beds, the area is also characterised by coralligenous assemblages and is a recognised nursery site for several juvenile of commercially-targeted species [75]. These expert opinions corroborate our findings and confirm the relevance of the established methodology for the selection of AAR deployment sites.

Conclusions

The comparison between the two pseudo-absence data generation methods reveals that the environmentally- and geographically-weighted method has significant potential to reduce prediction errors in SDMs. When confirmed absence data are not available, we recommend using the method proposed by Hengl et al. [37], which combines ENFA and RK predictions to deal with an imperfect ability to detect a species and incorporate spatial dependence in predictions. This study proposes a novel method to developing predicted maps that uses fuzzy sets to address the uncertainty associated with selecting probability threshold criteria. In the context of conservation planning, we illustrate the advantages of this novel method, using it to identify areas within the Gulf of Gabes that meet the conservation targets for AARs. Three key areas that met the conservation criteria to a high level were identified, and these areas are recommended as deployment

sites for AARs. The location and spatial arrangement of reef units must be carefully planned. Reef characteristics such as the number of modules, the distance between modules as a function of trawl parameters, and the weight of modules as a function of exposure to currents should all be informed by scientific studies. Additional factors that should be considered in the deployment of AARs are economic costs and socio-economic effects.

The areas with a nonzero favourability value for AAR deployment cover 1578 km^2 and encompass 30% of seagrass beds in good condition. However, these areas represent biodiversity targets without regard to cost. In this context this study can contribute to effective conservation planning in a broader prioritization process, which must be optimized by accounting for the cost of conservation. For instance, future works should focus on systematic conservation planning [76,77] that attempts to solve a cost-effectiveness problem (i.e. how to achieve the most conservation given limited resources). This is particularly important in the Gulf of Gabes where the diverse stakeholder group often holds conflicting values and opinions (e.g., conflicts between professional and local artisanal fishers).

Acknowledgments

Author Contributions

Conceived and designed the experiments: TH. Performed the experiments: TH. Analyzed the data: TH FLL FBRL. Contributed reagents/materials/analysis tools: TH CS MSR PC. Wrote the paper: TH FLL FBRL FL CA PC CS MSR.

References

1. Harris PT, Whiteway T (2009) High seas marine protected areas: Benthic environmental conservation priorities from a GIS analysis of global ocean biophysical data. Ocean & Coastal Management 52: 22–38. doi:10.1016/j.ocecoaman.2008.09.009.

2. Valavanis VD, Pierce GJ, Zuur AF, Palialexis A, et al (2008) Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia 612: 5–20. doi:10.1007/s10750-008-9493-y.

3. Johnson AF, Jenkins SR, Hiddink JG, Hinz H (2012) Linking temperate demersal fish species to habitat: scales, patterns and future directions. Fish and Fisheries. doi:10.1111/j.1467-2979.2012.00466.x.

4. Degraer S, Verfaillie E, Willems W, Adriaens E, Vincx M, et al (2008) Habitat suitability modelling as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea. Continental Shelf Research 28: 369–379. doi:10.1016/j.csr.2007.09.001.

5. Maxwell DL, Stelzenmüller V, Eastwood PD, Rogers SI (2009) Modelling the spatial distribution of plaice (Pleuronectes platessa), sole (Solea solea) and thornback ray (Raja clavata) in UK waters for marine management and planning. Journal of Sea Research 61: 258–267. doi:10.1016/j.seares.2008.11.008.

6. MacLeod CD, Mandleberg L, Schweder C, Bannon SM, Pierce GJ (2008) A comparison of approaches for modelling the occurrence of marine animals. Hydrobiologia 612: 21–32. doi:10.1007/s10750-008-9491-0.

7. Elith J, Leathwick JR (2009) Species Distribution Models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

8. Hoffman JC, Bonzek CF, Latour RJ (2009) Estimation of Bottom Trawl Catch Efficiency for Two Demersal Fishes, the Atlantic Croaker and White Perch, in Chesapeake Bay. Marine and Coastal Fisheries 1: 255–269. doi:10.1577/C08-048.1.

9. Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157: 89–100.

10. Anderson R. (2003) Real vs. artefactual absences in species distributions: tests for Oryzomys albigularis (Rodentia: Muridae) in Venezuela. Journal of Biogeography 30: 591–605.

11. Loiselle B., Howell C., Graham C., Goerck J., Brooks T, et al (2003) Evitando Dificultades Resultantes del Uso de Modelos de Distribución de Especies en Planeación de Conservación. Conservation Biology 17: 1591–1600. doi:10.1111/j.1523-1739.2003.00233.x.

12. Robinson L, Elith J, Hobday A, Pearson R, Kendall B, et al (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology and Biogeography 20: 789–802.

13. Monk J (2013) How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? Fish and Fisheries. doi:10.1111/faf.12039.

14. Mackenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42: 1105–1114. doi:10.1111/j.1365-2664.2005.01098.x.

15. Anderson R., Dudík M, Ferrier S, Guisan A, J Hijmans R, et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

16. Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83: 2027–2036.

17. Stockwell D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13: 143–158. doi:10.1080/136588199241391.

18. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

19. Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555–1568. doi:10.1111/j.1365-2699.2004.01076.x.

20. Elith J, H. Graham C, P. Anderson R, Dudík M, Ferrier S, et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

21. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3: 327–338. doi:10.1111/j.2041-210X.2011.00172.x.

22. Engler R, Guisan A, Rechsteiner L (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41: 263–274.

23. Torres LG, Read AJ, Halpin P (2008) Fine-scale habitat modeling of a top marine predator: do prey data improve predictive capacity. Ecological Applications 18: 1702–1717.

24. Jones MC, Dye SR, Pinnegar JK, Warren R, Cheung WW (2012) Modelling commercial fish distributions: Prediction and assessment using different approaches. Ecological Modelling 225: 133–145.

25. Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling 210: 478–486. doi:10.1016/j.ecolmodel.2007.08.010.

26. Wisz MS, Guisan A (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecology 9: 8. doi:10.1186/1472-6785-9-8.

27. Freeman EA, Moisen GG (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modelling 217: 48–58. doi:10.1016/j.ecolmodel.2008.05.015.

28. Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43: 393–404. doi:10.1111/j.1365-2664.2006.01149.x.

29. Albouy C, Guilhaumon F, Araújo MB, Mouillot D, Leprieur F (2012) Combining projected changes in species richness and composition reveals climate change impacts on coastal Mediterranean fish assemblages. Global Change Biology 18: 2995–3003. doi:10.1111/j.1365-2486.2012.02772.x.

30. Nenzén HK, Araújo M (2011) Choice of threshold alters projections of species range shifts under climate change. Ecological Modelling 222: 3346–3354.

31. Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122: 99–112. doi:10.1016/j.biocon.2004.07.004.

32. Teh LCL, Teh LSL (2011) A fuzzy logic approach to marine spatial management. Environmental Management 47: 536–545.

33. Montero J (2009) Fuzzy Logic and Science. In: Seising R, editor. Views on Fuzzy Sets and Systems from Different Perspectives. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Vol. 243. 67–77.

34. Hattour A (1991) Le chalutage dans les eaux Tunisiennes réalités et considérations législatives particulièrement dans les Golfe de Tunis et de Gabès. Note de l'Institut National Scientifique et Technique d'Océanographie et de Pêche de Salammbô: 13p.

35. Najar B, Ben Mariem S, Hadj Ali M (2010) Évolution des profils des débarquements de poissons dans la région de Gabes, Tunisie. Commission International pour l'Exploration Scientifique de la Mer Méditerranée 39: 601p.

36. Direction Générale de la Pêche et de l'Aquaculture (DGPA) (2010) Annuaire des statistiques des pêches en Tunisie. Ministère de l'Agriculture, Tunisie.

37. Hengl T, Sierdsema H, Radović A, Dilo A (2009) Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling 220: 3499–3511. doi:10.1016/j.ecolmodel.2009.06.038.

38. Batisse M, De Grissa AJ (1995) Marine Region 3: Mediterranean. In G Kelleher, C Bleakley and S Wells (eds) A Global Representative System of Marine Protected Areas Vol I, The Great Barrier Reef Marine Park Authority, The World Bank and IUCN, Washington, DC: 77–104.

39. Francour P (1997) Fish Assemblages of Posidonia oceanica Beds at Port-Cros (France, NW Mediterranean): Assessment of Composition and Long-Term Fluctuations by Visual Census. Marine Ecology 18: 157–173. doi:10.1111/j.1439-0485.1997.tb00434.x.

40. Ben Mustapha K, Hattour A, Mhetli M, El Abed A, Tritar B (1999) Etat de la Bionomie Benthique des Etages Infra et Circalittoral du Golfe de Gabès. Bulletin de l'Institut National des Sciences et Technologies de la Mer 26: 5–48.

41. Jensen A, Collins K, Lockwood AP (2000) Artificial Reefs in European Seas. Lockwood, A.P. (Eds.). Springer. 508 p.

42. Munoz-Perez JJ, Mas G, Jose M, Naranjo JM, Torres E, et al (2000) Position and monitoring of anti-trawling reefs in the Cape of Trafalgar (Gulf of Cadiz, SW Spain). Bulletin of Marine Science 67: 761–772.

43. Moore C., Harvey E., Niel K (2010) The application of predicted habitat models to investigate the spatial ecology of demersal fish assemblages. Marine Biology 157: 2717–2729. doi:10.1007/s00227-010-1531-4.

44. Monk J, Ierodiaconou D, Bellgrove A, Harvey E, Laurenson L (2011) Remotely sensed hydroacoustics and observation data for predicting fish habitat suitability. Continental Shelf Research 31: 17–27. doi:10.1016/j.csr.2010.02.012.

45. Katsanevakis S, Maravelias CD (2008) Bathymetric distribution of demersal fish in the Aegean and Ionian Seas based on generalized additive modeling. Fisheries Science 75: 13–23. doi:10.1007/s12562-008-0033-5.

46. Pittman SJ, Brown KA (2011) Multi-Scale Approach for Predicting Fish Species Distributions across Coral Reef Seascapes. PLoS ONE 6: e20583. doi:10.1371/journal.pone.0020583.

47. Palialexis A, Georgakarakos S, Karakassis I, Lika K, Valavanis VD (2011) Prediction of marine species distribution from presence–absence acoustic data: comparing the fitting efficiency and the predictive capacity of conventional and novel distribution models. Hydrobiologia 670: 241–266. doi:10.1007/s10750-011-0673-9.

48. Monk J, Ierodiaconou D, Harvey E, Rattray A, Versace VL (2012) Are We Predicting the Actual or Apparent Distribution of Temperate Marine Fishes? PLoS ONE 7: e34558. doi:10.1371/journal.pone.0034558.

49. Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A, Hortal J (2008) Challenging species distribution models: the case of Maculinea nausithous in the Iberian Peninsula. Annales Zoologici Fennici. Vol. 45. 200–210.

50. Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant E, Veran S (2012) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution: 236–243. doi:10.1111/2041-210x12004.

51. Monk J, Ierodiaconou D, Versace VL, Bellgrove A, Harvey E, et al (2010) Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. Marine Ecology Progress Series 420: 157–174.

52. Calenge C (2006) The package "adehabitat" for the R software: A tool for the analysis of space and habitat use by animals. Ecological Modelling 197: 516–519. doi:10.1016/j.ecolmodel.2006.03.017.

53. R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria. Available: http://www.R-project.org/.

54. Montgomery DC (2007) Design and Analysis of Experiments, 6th Edition Set. New York: Wiley. 752 p.

55. Hill M, Smith A (1976) Principal component analysis of taxonomic data with multi-state discrete characters. Taxon 25: 249–255.

56. McCullagh P, Nelder JA (1989) Generalized Linear Models, Second Edition. 2nd edition London. Chapman and Hall. 512 p.

57. Miller J, Franklin J, Aspinall R (2007) Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202: 225–242. doi:10.1016/j.ecolmodel.2006.12.012.

58. Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: From equations to case studies. Computers & Geosciences 33: 1301–1315. doi:10.1016/j.cageo.2007.05.001.

59. Chambers JM, Hastie T, editors (1991) Statistical models in S. London: Chapman & Hall. 608 p.

60. Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence-absence models. Environmental Conservation 24: 38–49.

61. Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17: 145–151.

62. Austin M (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200: 1–19. doi:10.1016/j.ecolmodel.2006.07.005.

63. Zheng B, Agresti A (2000) Summarizing the predictive power of a generalized linear model. Statistics in medicine 19: 1771–1781.

64. Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38: 921–931.

65. Zadeh L (1965) Fuzzy sets. Information and Control 8: 338–353. doi:10.1016/S0019-9958(65)90241-X.

66. Hanberry B., He H., Palik BJ (2012) Pseudo-absence generation strategies for Species Distribution Models. PloS one 7: e44486.

67. Zarnetske PL, Edwards Jr TC, Moisen GG (2007) Habitat classification modeling with incomplete data: Pushing the habitat envelope. Ecological Applications 17: 1714–1726.

68. King G, Zeng L (2001) Logistic Regression in Rare Events Data. Political Analysis 9: 137–163.

69. Dixon PM, Ellison AM, Gotelli NJ (2005) Improving the precision of estimates of the frequency of rare events. Ecology 86: 1114–1123. doi:10.1890/04-0601.

70. McPherson JM, Jetz W, Rogers D (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Journal of Applied Ecology 41: 811–823. doi:10.1111/j.0021-8901.2004.00943.x.

71. Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385–393. doi:10.1111/j.0906-7590.2005.03957.x.

72. Williams AK (2003) The influence of probability of detection when modeling species occurrence using GIS and survey data [PhD thesis]. Virginia Polytechnic Institute and State University. Available: http://vtechworks.lib.vt.edu/handle/10919/11129. Accessed 1 October 2012.

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

74. El Lakhrach H, Hattour A, Jarboui O, Elhasni K, Ramos-Espla A (2012) Spatial distribution and abundance of the megabenthic fauna community in Gabes gulf (Tunisia, eastern Mediterranean Sea). Mediterranean Marine Science 13: 12–29.

75. Ben Mustapha K, Afli A (2007) Quelques traits de la biodiversité marine de Tunisie: Proposition d'aires de conservation et de gestion. Report of the MedSudMed Expert Consultation on Marine Protected Areas and Fisheries Management. MedSudMed Technical Documents. Rome (Italy). 32–55.

76. Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405: 243–253.

77. Watson JE, Grantham H, Wilson KA, Possingham HP (2011) Systematic conservation planning: Past, present and future. Wiley-Blackwell: 136–160.