NOO3D: A procedure to perform 3D species distribution models

Elisa Pérez-Costas[a,b], Cástor Guisande[a,*], Luis González-Vilas[a], Emilio García-Roselló[c], Jürgen Heine[c], Jacinto González-Dacosta[c], Jorge M. Lobo[c]

[a] Faculty of Science, Universidad de Vigo, 36310 Vigo, Spain
[b] Computer Science Department, Fundición Building, 26310, Universidad de Vigo, Vigo, Spain
[c] Biogeography and Global Change Departament, Museo Nacional de Ciencias Naturales (CSIC), c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

ABSTRACT

There is consensus surrounding the need to include a third dimension when estimating Species Distribution Models (SDMs), which is of special interest for marine species. Application of the third dimension is, however, rarely available, thus users are obliged to manually combine 2D SDM outputs (i.e., suitability or presence/absence maps) for 3D distribution generation. Herein, the Niche of Occurrence 3D (NOO3D) is presented, which is a new, simple modelling procedure that provides 3D distributions using both 3D occurrence samples and environmental datasets that consist of one layer per depth value. NOO3D performance was evaluated using five virtual marine species to avoid errors associated with real data sets (three pelagic species, with wide, medium, and narrow distributions, respectively, a mesopelagic species and an abyssal species). These virtual species are distributed across the North Atlantic Ocean and were built to a 0.5° x 0.5° resolution and considering 49 depth levels (from 0.43 m to an undersea depth of 5274.7 m). NOO3D results were also compared to those provided by 3D Alpha Shapes and Maximum Entropy (MaxEnt). The True Positive Rate (TPR), or sensitivity, True Negative Rate (TNR), or specificity, False Positive Rate (FPR), or commission error, and False Negative Rate (FNR), or omission error, were employed in order to facilitate comparison between methods. MaxEnt performed best for TPR, TSS and FNR, and Alpha Shape 3D performed best for FPR and TNR. NOO3D was always the second-ranked method for all metrics considered, which indicates that it was the most suitable method. The provided results indicate that NOO3D can be considered a viable alternative in achieving three-dimensional species distribution models.

1. Introduction

Species Distribution Models (SDMs), which are also known in the literature as ecological niche models, bioclimatic models, or habitat-suitability models, are procedures based on different statistical approaches, that relate species occurrence data to environmental conditions and aim to estimate suitable localities (Guisan et al., 2017). Due to the increase in computer processing power and the broader availability of digital environmental layers and species distribution data, SDM use has grown exponentially in recent years in ecological, biodiversity, and conservation studies (Hortal et al., 2012). However, both SDM results and accuracy are heavily influenced by certain limitations, which include the quality and biases of the occurrence data used (Amboni and Laffan, 2012), the frequent utilization of arbitrary areas in the model-building process (Barve et al., 2011), and the lack of reliable absence data (Lobo et al., 2018). Consequently, SDMs often rely on pseudo or background absences that are randomly selected from the study area, a procedure that tends to generate misleading predictions and erroneous validation results (Aarts et al., 2012; Hastie and Fithian, 2013; Hijmans, 2012; Lobo et al., 2008).

SDM extension to three dimensions has been advocated (Duffy and Chown, 2017), mainly for marine pelagic species. However, the main impediments to following this recommendation are the lack of information regarding the three-dimensional (3D) occurrence of most marine species and the dominance of records from shallow waters (Webb et al., 2010). Many different procedures have been used to predict marine species distributions (see Dambach and Rödder, 2011; Redfern et al., 2006; Robinson et al., 2011; Valavanis et al., 2008), but few estimate 3D distributions. When approached, this objective was confronted by stacking 2D SDMs from different depths (Bentlage et al., 2013). These were implemented using simple, but conceptually sound procedures, based on the fact that there is limited reliable marine species data (Cheung et al., 2007; Kaschner et al., 2006), or through the use of complex modelling techniques that apply presence-background

---

absence approaches (Alabia et al., 2016; Bentlage et al., 2013). Given that 3D SDMs are based upon the simple combination of 2D SDM outputs (i.e., suitability or presence/absence maps), the aim here was to introduce a new, simple modelling procedure that would be able to provide 3D distributions for applied of basic purposes. This procedure is based in a simple method directed to estimate the probable distribution of species called Niche of Occurrence (NOO; see García-Roselló et al., 2019) but now considering the complete three-dimensional gradient of environmental conditions corresponding to a marine environment (NOO3D). The main features of this procedure are explained initially, in order to subsequently compare its performance with other simple geometrical procedures, such as 3D Alpha Shapes (Lafarge and Pateiro-Lopez, 2017), and also with one of the most widely-used applications in the SDM field (MaxEnt; Phillips et al., 2006). The purpose of these comparisons is not to demonstrate the lack of reliability of the other modelling procedures, but only to show the potentialities and performance capacity of the proposed application.

2. Methods

2.1. The niche of occurrence (NOO) approach

Both NOO and NOO3D are available in ModestR software, a free application for the management of taxonomic data and construction of species distribution predictions (García-Roselló et al., 2013, 2014). NOO3D is based on a simple method called NOO (see García-Roselló et al., 2019) which is applied to estimate the probable distribution of species using exclusively the available presence data. A detailed description of the procedure is provided in García-Roselló et al. (2019) and in the step-by-step tutorials freely available at http://www.ipez.es/modestr/Manual_Tutorial.html. Next, we will briefly describe the basic steps of the proposed procedure.

The approach begins with the delimitation of a Geographical Extent (GE) which defines the most probable target species accessible area. GE may be delimited using methods founded on the geometry of available occurrences, such as the convex hull, alpha shapes, or Kernel density distribution. As these delimitations may not strictly reflect natural units related to the accessible area, users may also select a predefined GE, in accordance with their a priori knowledge of the target species, or, for example, that of drainage basins or ecogeographic provinces.

The second step in our proposed procedure is the selection of the preferred environmental predictor variables, within the formerly predefined GE. To do so, a collinearity analysis is performed through the computation of the Variance Inflation Factor (VIF) (Fox and Weisberg, 2011), so that redundant variables with VIF values above a predefined threshold are removed. Variables are considered to be highly correlated when their VIF values are above 30 (default ModestR value), however this threshold may be modified by users at their convenience. The contribution of the previously selected variables for target species distribution is then analysed, using the Instability Index approach (Guisande, 2019; Guisande et al., 2017). This method aims to identify predictors that best differentiate between those environmental conditions which prevail at presence locations and those available in the GE area. Users may include only variables with high contributions to this index by selecting an accumulated contribution percentage (default value is 80%), such that all variables are included, if 100% is selected.

Once preferred environmental predictor variables have been selected the next step is the generation of a Compounded Environmental Layer (CEL) using polar coordinates (Van Sickle, 2010). Polar coordinates are useful for the representation of any number of different variables in two-dimensional coordinate systems through the assignment of an angle to each variable, i.e. a polar coordinate system axis, and a distance (positive or negative values) from the centre. NOO assigns an angle to each selected environmental variable by calculating the correlation matrix among them, and establishing a default order, such that each variable is followed by that which is most highly

correlated. This default order may also be modified by users. Distances are computed following the standardization of environmental variables, from 1 to 2, in each cell of the considered GE. A CEL may be displayed in ModestR either in polar space or geographically, by translating the values of the polar coordinate system to a map. However, its main purpose is to obtain and save all information about relevant predictors related to species environmental preferences in the selected GE, in a single layer.

Subsequently, target species occurrence cells are projected onto the CEL, and a kernel density estimation is computed to reflect the intensity of these presences in the environmental space. The minimum density value attained in a presence cell is selected as the default threshold (minimal density at presence option), in order to define the species' suitable area. Consequently, only those cells with environmental conditions similar to those which exist in the occurrence cells are selected as suitable. Finally, density in the environmental polar space is projected geographically, so as to obtain the final output, which may either be a binary map that shows suitable and unsuitable cells within the GE (distribution map option), or a continuous map (density map option) that reflects the quantity of presence observations for each set of environmental conditions. Only two parameters must be selected by users in this final step: a smoothing factor and a tolerance value. The smoothing factor is the bandwidth or deviation of the smoothing kernel, such that greater values produce a greater spread around presence cells. This tolerance allows for the expansion of minimum and maximum values of the selected environmental variables, found in the presence cells. In accordance with previous virtual species tests, a smoothing factor of two and a tolerance of 1% provide generally adequate results, and are thus defined as default values in ModestR (see García-Roselló et al., 2019).

2.2. NOO3D

NOO3D extends the usage of NOO (García-Roselló et al., 2019) to a 3D space, using both environmental variables and presence cells that include a third, Z dimension (in our case, the vertical distribution in the 3D marine space). This is a custom dimension defined by users. However, in the present case, it is the depth of marine environments. Complete details on the procedure and the different options available are explained in the Appendix and in the tutorial 19 freely available at http://www.ipez.es/modestr/Manual_Tutorial.html, illustrated by using a real example with the data of the Atlantic cod (*Gadus morphua*), a pelagic species with an extended geographical distribution across a wide range of depths.

As in NOO, the first step in NOO3D is the selection of a Geographical Extent (GE), by following similar procedures, which are delimited through consideration of the depth range of the target species presences. The second step, again, is the selection of preferred environmental variables. Unlike NOO, NOO3D uses 3D structured variables which consist of a set of 2D environmental layers at different Z values (i.e. considering the complete set of 3D values). Both collinearity analysis, based on the Variance Inflation Factor (VIF), and contribution analysis, using the above-mentioned Instability Index are available for this step to select the most important predictors on the basis of the complete values of environmental variables in the entirety of the 3D space. Instead of using individual 2D environmental layers and models carried out at different depths as in the case of Bentlage et al. (2013), the proposed third step is the generation of a 3D compounded environmental layer (3D CEL), which is composed by the set of 2D compounded environmental layers (2D CEL; i.e. one 2D CEL at each Z level). This step cannot be considered an advantage directed to obtain better model results, but only a procedure able to compile all the environmental information in a unique file. One important consideration here is that the angles and distances of each one of these 2D CELs are computed using all cells throughout the complete GE, including values from all Z levels. The default 3D CEL order is also established, as a

function of the correlation matrix, among selected 3D variables.

Once the 3D CEL has been built and stored, the values of the selected environmental variables for each occurrence cell is obtained and represented in its corresponding 2D CEL, in order to estimate a kernel density at each Z level. This kernel estimation represents the intensity of these presences in the polar space. As in NOO, the suitable species area is defined in the environmental polar space as the minimum density attained in a presence cell (minimal density at presence option). Finally, density in the environmental polar space is geographically projected, in order to obtain a binary (distribution map option) or continuous suitability (density map option) map, for each Z level.

2.3. Input variables employed

Marine environmental data were obtained, in January of 2018 from the COPERNICUS website (http://marine.copernicus.eu/), a European program which provides the scientific community with various data products for marine applications. The COPERNICUS website facilities allowed for the selection of six variables, which were downloaded as netCDF files: two physical variables (water temperature and salinity) and four biogeochemical variables (chlorophyll, phosphate, silicate, and dissolved iron concentrations). The temporal coverage of physical and biogeochemical variables is similar (from 01/16/2012 to 12/16/2017 and from 01/14/2012 to 12/16/2017, respectively). With the tools available in ModestR, the complete period for each variable and depth level was averaged, and the resulting layers were then rescaled to a common geographical grid of 0.5° x 0.5°, so as to correct differences in spatial resolution between physical and biogeochemical variables (0.083° x 0.083 and 0.5° x 0.5°, respectively). Each variable was finally grouped as a 3D environmental layer, composed of a raster file, for each of the 49 depth levels considered, from 0.43 m to an undersea depth of 5274.7 m.

2.4. Building virtual species

NOO3D performance was assessed by comparing the predictions generated with those of virtual marine species, whose complete distributions were simulated. For this, the *virtualspecies* R package, developed by Leroy et al. (2016), was utilized. Thus, the 3D distributions of five virtual marine species were built in the North Atlantic Ocean (FAO areas 21 and 27): three pelagic species, with wide, medium, and narrow distributions, respectively, a mesopelagic species and an abyssal species. Depth ranges are shown in Table 1. This package provides two different approaches (response and PCA) to build random suitability maps from a spatial dataset (raster stack) of environmental layers. The response method, which is based on the application of response functions to each environmental variable and combining the results, in order to generate final environmental suitability maps, was selected. Virtual species were based only on three 3D environmental variables, phosphate concentration, water temperature, and salinity, which were randomly selected from the six mentioned above. These three variables were cropped to the study area, i.e., the North Atlantic Ocean (FAO areas 21 and 27), and masked ice-covered areas and the Baltic Sea

Table 1
Summary of the used virtual species. The number of depth levels (DL), depth range (DR, in m), the total number of considered cells in the geographical extent(C),the total number of presence cells used in the modelling process (PC), and real prevalence of the virtual species (P) are shown.

	DL	DR	C	PC	P
Pelagic low	26	0.5–186.1	240,668	23,713	0.10
Pelagic medium				113,636	0.47
Pelagic high				217,112	0.90
Mesopelagic	9	222.5–902.3	55,745	28,244	0.51
Abyssal	8	1062.4–2865.5	35,716	15,743	0.44

(owing to its low salinity).

For the pelagic species, multiple random virtual distributions were generated in a loop, using only the water surface level of the three selected variables and a multiplicative combination of three types of response functions (Gaussian, linear and logistic). Each response was scaled, and the final probability map reflected levels between 0 and 1. For each virtual distribution, the suitability map generated was converted into a presence/absence map, in accordance with three different species prevalence values: 0.1 (narrow distribution), 0.5 (medium distribution) and 0.9 (wide distribution) (Table 1). Different prevalence values were used because they may influence model results (Fukuda and De Baets, 2016) and we aim to estimate the general performance of the proposed method under different situations. The complete 3D distribution of each species was obtained by applying the same input parameters as the surface level (i.e., the same response functions for each variable) for all depth levels considered, in accordance with the values of those environmental variables employed.

The same methodology was applied for obtaining both the mesopelagic and abyssal 3D distribution of virtual species, although instead of using the surface level as a reference to obtain virtual distribution parameters, data corresponding to the undersea depth of 541 m was utilized for the mesopelagic species, and the undersea depth of 1245 m was applied for the abyssal species. Moreover, in these cases, a distribution was chosen, for each case, with a prevalence of approximately 0.5 (medium distribution) (Table 1).

2.5. Modelling methods

Species distribution predictions were carried out using all six physical and biogeochemical variables as predictors, not just those three which had been selected in virtual species generation. This was done intentionally, in order to simulate the modelling process, under real conditions, in which true causal variables are unknown. Note that all of these 3D variables were already cropped to the study area, and were also limited to the corresponding depth range.

NOO3D were developed using only verified 3D presence cells, as defined by their longitude, latitude, and depth. Each presence cell was extracted through the random selection of a depth level, followed by the selection of cells on the corresponding virtual species suitability map. For each one of the five considered virtual species, five percentages (very low = 0.1%; low = 1%, medium = 5%, high = 10% and very high = 20%) for the total number of presence cells (including all depth levels) were randomly selected, and the process was repeated 10 times, in order to generate a total of 50 presence sets. Thus, a total of 250 presence sets (5 virtual species x 5 presence percentages x 10 repetitions) were finally processed, using NOO3D. In the generation of the 3DCEL, a more restrictive 10 VIF threshold and an accumulated percentage of contribution of 100%, for the instability index, were used. Output continuous suitability maps were generated using a smooth factor of two, and a tolerance value of 0.1%.

During the last decade, maximum entropy (MaxEnt, Phillips et al., 2006) has been the most widely used algorithm to predict species distribution from partial data. MaxEnt is a machine learning method that uses the maximum entropy principle to estimate the probable geographical distribution of a species by generally distinguishing between presence and background environmental conditions. Numerous authors have provided an exhaustive analysis of its foundations and/or performance in different scenarios (Elith et al., 2011). It requires a large number of background locations that are often randomly sampled from non-presence cells, because information about species absences is rarely provided, or is unreliable, due to the lack of survey effort information (Lobo et al., 2010). As in most situations, MaxEnt has been applied herein using presences and background absences. For each one of the 250 presence sets used in NOO3D, a background absence set was created by selecting background (non-presence) cells at random from the 3D geographical extent. As proposed by García-Roselló et al. (2019),
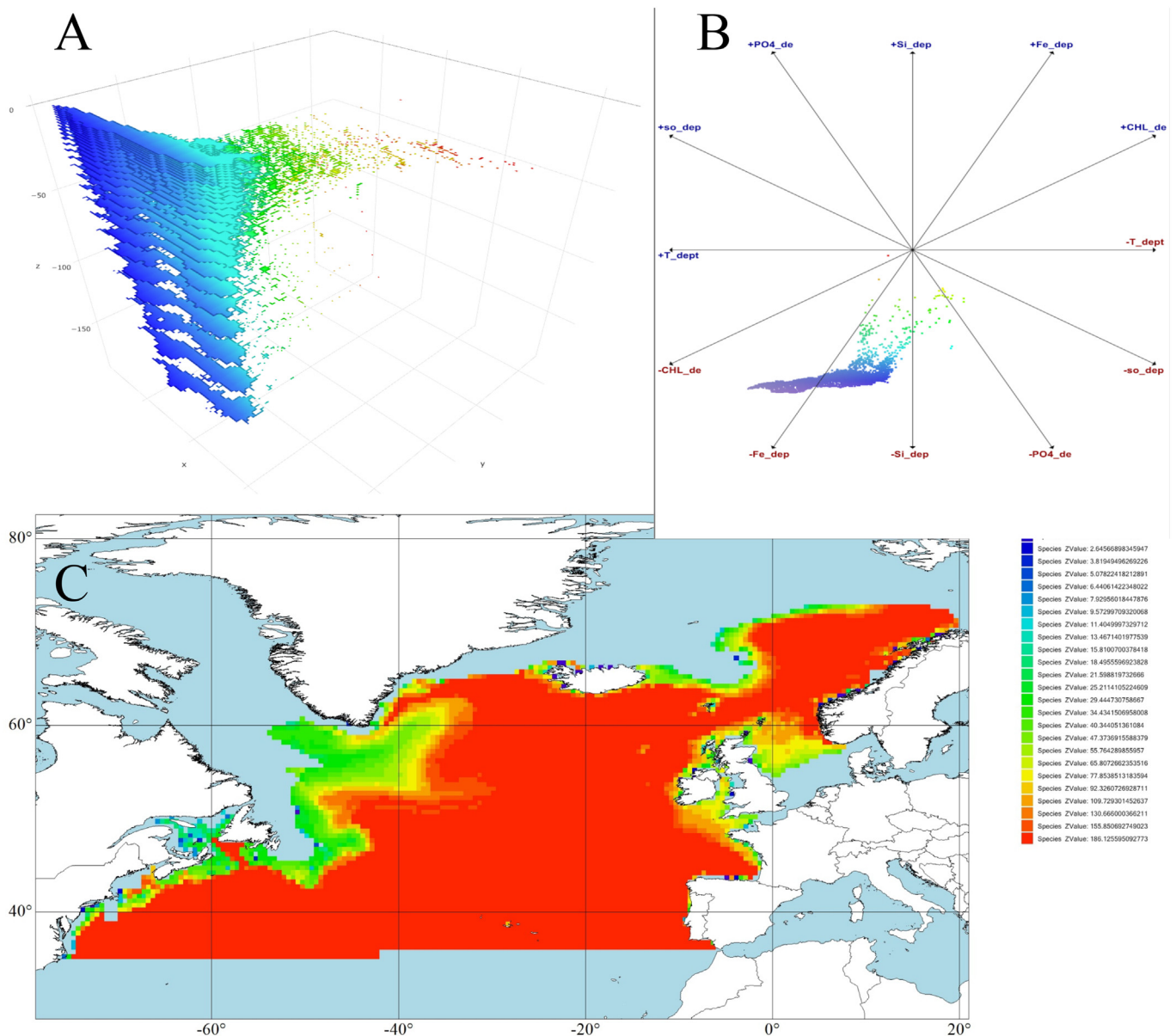
Fig. 1. A) 3D plot of the 3D compounded environmental layer (3D CEL) used for the pelagic species, including 26 depth levels ranged from 0 to 200 m. Z axis shows depth, while X and Y axes are the polar coordinates. B) Plot of polar coordinates showing the environmental layer combining all the variables (temperature, salinity, phosphate, silicate, iron and chlorophyll) for the surface level (0.4 m). C) Distribution map of the pelagic medium virtual species. Colour legend shows the maximum depth at which the species is present.

the number of background absences was set to 5% of total cells in the complete study region in order to simulate situations in which true absences are unavailable. The *dismo* R package was used for this purpose. 3D distributions (i.e., a continuous suitability map for each depth layer) were obtained through the use of two different methods for each of the 250 presence/background absence sets: MaxEnt all layers and MaxEnt by layer. In both cases, MaxEnt models are built using the same environmental predictors as used in NOO3D and default options. MaxEnt all layers creates a unique model using environmental data linked with the complete presence/background absence set, and hence including data from all the depth levels. This model is then applied to each depth level in order to obtain the corresponding continuous suitability maps, which are eventually included in the complete 3D distribution. Conversely, MaxEnt by layer generates a different model for each depth level using only environmental data derived from presences and background absences at that specific depth level. Each model is then applied to build the corresponding suitability maps making the

complete 3D distribution. Some authors have found that selecting different model parameters influencing the complexity of the established relationships may vary the results of MaxEnt when dealing with a low number of presences (see Shcheglovitova and Anderson, 2013; Warren and Seifert, 2011). As the use of background absences prevents the assessment of model accuracy (Lobo et al., 2008), the so obtained distributions may vary considerably in response to different settings (see Merow et al., 2013). Furthermore, the criterion used to select MaxEnt settings does not allow generating geographical predictions with low commission and omission errors (Velasco and González-Salazar, 2019). Due to all these reasons, we have preferred to maintain default selection criteria in order to avoid getting output geographical distributions similar to the ones used to feed the model.

The alphashape3D R package (Lafarge and Pateiro-Lopez, 2017) was also used to provide species predictions, based only on occurrence data geometry (Alpha Shape 3D). This method uses a finite set of points in a three-dimensional space, and was applied here to each of the 250
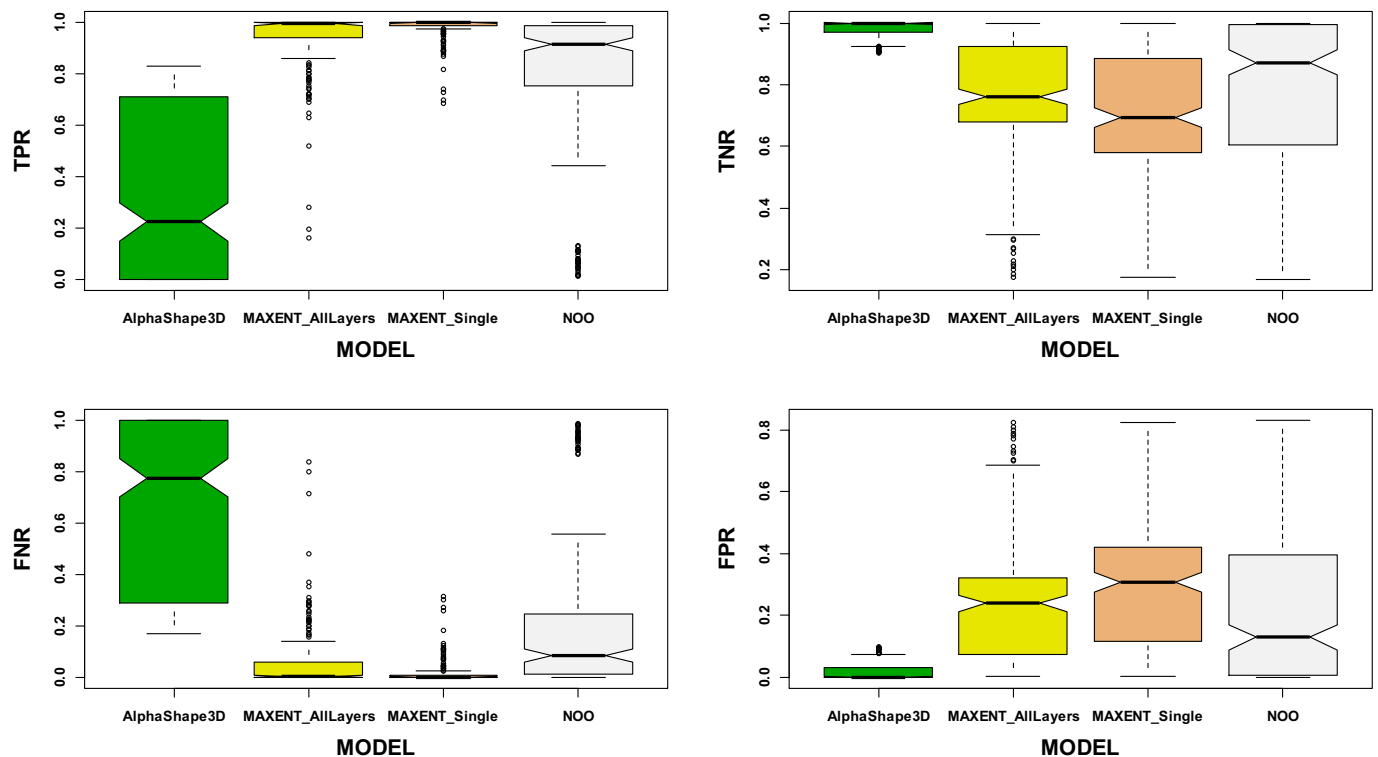
Fig. 2. Notched box plots which represent the variation in the performance metrics used for the four modelling methods considered. TPR = True Positive Rate, or sensitivity, TNR = True Negative Rate, or specificity, FPR = False Positive Rate, or commission errors, FNR = False Negative Rate, or omission errors. Notched box plots show median values (horizontal line), interquartile range values between upper and lower quartiles (top and bottom of the box), distribution of 99% of data (upper and lower dashed lines), outliers (open circles), and notch lengths representing classic 95% confidence intervals, such that, when notches do not overlap, medians may be seen to differ significantly (Krzywinski and Altman, 2014).
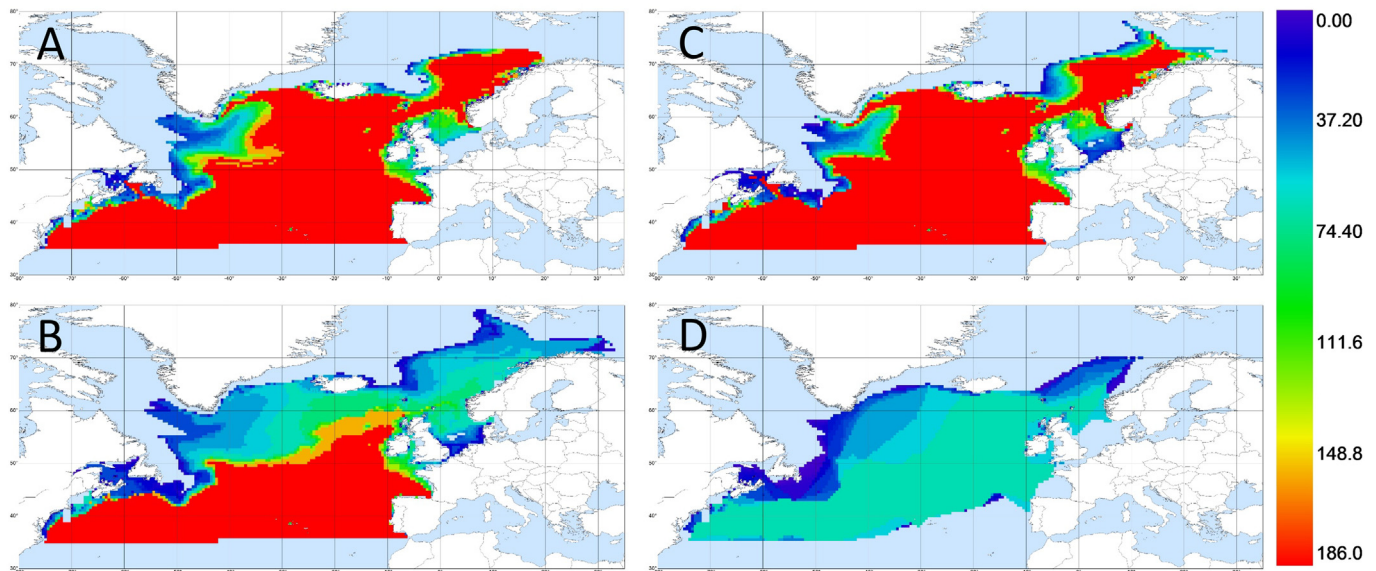

Fig. 3. Distribution map of the pelagic medium virtual species of the different models. Colour legend show the maximum depth at which the species is present. A) Representation of the NOO3D Model, B) MaxEnt All Layers model, C) MaxEnt Single Layer model, and D) AlphaShape 3D model.

presence sets, so as to obtain a 3D distribution, i.e., a binary presence/absence map for each depth level within the range defined for the corresponding virtual species.

2.6. Model validation

SDM model performance is usually based on a set of discrimination measures which compare predictions to a more-or-less independent

validation presence-absence dataset (Fielding and Bell, 1997). The vast of the these metrics come from a confusion matrix, which is built by comparing this validation presence-absence dataset to model binary output. In the case of models with a continuous suitability output (e.g. MaxEnt), a binary output must be obtained by the initial application of a predefined threshold (Lobo et al., 2008). In this particular case, with both the MaxEnt all layers and in MaxEnt by layer options, presence/absence maps for each depth level were obtained using the Minimum

Training Presence Threshold (MTPT), or the lowest predicted value associated with an observed presence. The main four performance metrics derived from the confusion matrix were used herein to compare obtained predictions with the "true" map of each virtual species: the True Positive Rate (TPR), or sensitivity, True Negative Rate (TNR), or specificity, False Positive Rate (FPR), or commission error, and False Negative Rate (FNR), or omission error. These four metrics are selected because they constitute the basis for all the discrimination performance measures (Fielding and Bell, 1997).

2.7. Statistical analyses and graphics

Model results were compared using a graphic tool, based on statistics that do not require assumptions about the shape of the distribution of considered values. In order to do this, notched box plots which represented Median values (M), Interquartile Ranges (IQR = Q3-Q1 representing 50% of the data), and notch lengths (M $\pm$ 1.58 x IQR/√n) were employed. In these boxplots, notches approximated classic 95% confidence intervals, so that when notches did not overlap, medians could be seen to differ significantly (Krzywinski and Altman, 2014) without the requirement of additional statistical analyses (Chambers et al., 1983). All of these analyses were performed with the StatR and PlostR RWizard applications (Guisande et al., 2014).

3. Results

Fig. 1 shows an example of the 3D plots that may be obtained from ModestR, which enable the display of three-dimensional variation in the number of occurrences (Fig. 1A), as well as their distribution along a two-dimensional polar coordinate system (Fig. 1B) or throughout the geographical space (Fig. 1C).

The sensitivity (correctly predicted presences or true positive rate) was high and similar in the two MaxEnt modelling procedures, although MaxEnt by layer showed less dispersion around the median. Sensitivity was significantly lower in NOO3D (notches did not overlap), and was very poor in Alpha Shape 3D (Fig. 2). This result matches the higher values for false negative rates (omission errors) in NOO3D (median = 0.21), compared to MaxEnt results. However, specificity (absences correctly predicted or true negative rate) was significantly higher in NOO3D than in the two MaxEnt methods, and was even higher in Alpha Shape 3D, with values near zero and with low dispersion around the median value (Fig. 2). Consequently, commission errors (true positive rates) were significantly lower in NOO3D than in MaxEnt predictions.

Fig. 3 shows the results obtained for a pelagic medium virtual species using the three methods, as an example for comparing the four models. NOO3D is more similar to MaxEnt Single Layer model, than to the other models: MaxEnt All Layers model and AlphaShape 3D model.

4. Discussion

We consider that a study such as this should not go beyond the description of general results. Many factors may influence the obtained results when one aims to examine the specific results provided at different depths (the characteristics of the virtual species, the selected predictors, the species prevalence, the percentage of used presences, etc.) being risky to derive very specific recommendations. Thus, considering the usual conditions in which SDMs are applied (lack of reliable absence data and use of background absences, lack of delimitation of the accessible area, lack of information about species prevalence, and lack of information about causal predictors) the results provided demonstrate that the simple proposed approach (NOO3D) allows obtaining plausible distributional predictions. The use of different prevalence values, sampling sizes, and virtual species types indicate, by way of these results, that the procedure followed in NOO3D yields improved estimations regarding the places where species do not occur

(fewer commission errors), at the expense of inferior presence locality predictive capacity (increased omission errors).

The previous delimitation of the geographical extent for which each model was accomplished and projected, in the NOO3D procedure, could be a decisive option to wrongly predict those presences which occur outside of the selected area. Of course, this unavoidable weakness will be reduced when a high number presences are used (Stockwell and Peterson, 2002), and also when the used data are less biased both environmentally and geographically (El-Gabbas and Dormann, 2018; Sánchez-Fernández et al., 2011). The risk of leaving "true" occurrence localities outside the predicted distribution can be costly in some circumstances (e.g. in conservation planning), representing "missed opportunities" (Tulloch et al., 2016) that require additional and improved surveys capable of obtaining more representative data. However, this drawback is compensated in NOO3D by the advantage of including a significantly lower number of false presences in the localities in which the species are truly absent. A low rate of commission error is a valuable feature when distribution predictions are used for conservation purposes (Fielding and Bell, 1997; Rondinini et al., 2006) or to estimate species richness or compositional patterns from stacking individual SDMs (Ferrier and Guisan, 2006). Thus, the risk of selecting uninhabited, redundant or conservation ineffective areas is minimized when NOO3D is used.

The modelling procedure introduced here is intended for those occasions on which the user can only rely on presence information, which rules out the use of "false" or background absences (Lobo, 2016). The classic use-availability approach (Johnson, 1980) may be adequate for estimation of species environmental preferences, but not for generation of reliable predictions (Hastie and Fithian, 2013). NOO3D also forces the user to select a geographical extent, as well as the most probable predictors able to explain species preferences in the inhabited geographical area. These selections seem to improve the conceptual requirements of a distributional prediction, as the effects of causal or "semi-causal" variables must be inferred within the accessible area of each species. A final feature of the proposed application is that both the identity and comparative weight of the used explanatory variables are based on the values provided by the complete selected three-dimensional space. A combination of 2D SDMs which come from different depth ranges may be advantageous, if different predictors influence distribution at different depths (Dambach and Rödder, 2011). However, the dynamic 3D character of marine environmental variables and the decrease in spatial variation of environmental variables with depth (Costello et al., 2018) should promote modelling strategies that can cope with complete 3D gradients (including water pressure). Further improvement in 3D SDMs should be researched and implemented, so as to make predictions, taking into account the three-dimensional autocorrelated structure of marine variables and species (Washington and Parkinson, 2005).

References

Aarts, G., Fieberg, J., Matthiopoulos, J., 2012. Comparative interpretation of count, presence-absence and point methods for species distribution models. Methods Ecol. Evol. 3, 177–187.

Alabia, I.D., Saitoh, S.-I., Igarashi, H., Ishikawa, Y., Usui, N., Kamachi, M., Awaji, T., Seito, M., 2016. Ensemble squid habitat model using three-dimensional ocean data. ICES J. Mar. Sci. 73, 1863–1874.

Amboni, M.P., Laffan, S.W., 2012. The effect of species geographical distribution estimation methods on richness and phylogenetic diversity estimates. Int. J. Geogr. Inf. Sci. 26, 2097–2109.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol. Model. 222, 1810–1819.

Bentlage, B., Peterson, A.T., Barve, N., Cartwright, P., 2013. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. Glob. Ecol. Biogeogr. 22, 952–961.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 1983. Graphical Methods for Data Analysis. Wadsworth & Brooks/Cole.

Cheung, W.W.L., Watson, R., Morato, T., Pitcher, T.J., Pauly, D., 2007. Intrinsic vulnerability in the global fish catch. Mar. Ecol. Prog. Ser. 333, 1–12.

Costello, M.J., Basher, Z., Sayre, R., Breyer, S., Wright, D.J., 2018. Stratifying Ocean sampling globally and with depth to account for environmental variability. Sci. Rep. 8, 11259.

Dambach, J., Rödder, D., 2011. Applications and future challenges in marine species distribution modeling. Aquat. Conserv. 21, 92–100.

Duffy, G.A., Chown, S.L., 2017. Explicitly integrating a third dimension in marine species distribution modelling. Mar. Ecol. Prog. Ser. 564, 18.

El-Gabbas, A., Dormann, C.F., 2018. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. Ecol. Evol. 8, 2196–2206.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. J. Appl. Ecol. 43, 393–404.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, 2nd edition. Sage, Thousand Oaks, CA.

Fukuda, S., De Baets, B., 2016. Data prevalence matters when assessing species' responses using data-driven species distribution models. Ecol. Inform. 32, 69–78.

García-Roselló, E., Guisande, C., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., Vaamonde, A., Granado-Lorencio, C., 2013. ModestR: a software tool for managing and analyzing species distribution map databases. Ecography 36, 1202–1207.

García-Roselló, E., Guisande, C., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., González-Vilas, L., González-Dacosta, J., Vaamonde, A., Granado-Lorencio, C., 2014. Using ModestR to download, import and clean species distribution records. Methods Ecol. Evol. 5, 703–713.

García-Roselló, E., Guisande, C., González-Vilas, L., González-Dacosta, J., Heine, J., Pérez-Costas, E., Lobo, J.M., 2019. A simple method to estimate the probable distribution of species. Ecography 42, 1613–1622.

Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models, with Applications in R. Cambridge University Press, Cambridge.

Guisande, C., 2019. Estimation of the relative importance of factors affecting species distribution based on stability concept. R package version 1.8. In: CRAN.R-project.org/package=SPEDInstabR Version (01/2019).

Guisande, C., Heine, J., González-Dacosta, J., García-Roselló, E., 2014. RWizard Software. University of Vigo, Vigo, Spain.

Guisande, C., García-Roselló, E., Heine, J., González-Dacosta, J., González Vilas, L., García Pérez, B.J., Lobo, J.M., 2017. SPEDInstabR: an algorithm based on a fluctuation index for selecting predictors in species distribution modelling. Ecol. Inform. 37, 18–23.

Hastie, T., Fithian, W., 2013. Inference from presence-only data; the ongoing controversy. Ecography 36, 864–867.

Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93, 679–688.

Hortal, J., Lobo, J.M., Jiménez-Valverde, A., 2012. Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. Nat. Conserv. 10, 108–118.

Johnson, D.H., 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61, 65–71.

Kaschner, K., Watson, R., Trites, A.W., Pauly, D., 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Mar. Ecol. Prog. Ser. 316, 285–310.

Krzywinski, M., Altman, N., 2014. Visualizing samples with box plots: use box plots to illustrate the spread and differences of samples. Nat. Methods 11, 119–120.

Lafarge, T., Pateiro-Lopez, B., 2017. Implementation of the 3D alpha-shape for the reconstruction of 3D sets from a point cloud. R package version 1.3. CRAN.R-project.org/package=alphashape3d. In: Version (12/2017).

Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. Ecography 39, 599–607.

Lobo, J.M., 2016. The use of occurrence data to predict the effects of climate change on insects. Curr. Opin. Insect. Sci. 17, 62–68.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33, 103–114.

Lobo, J.M., Hortal, J., Yela, J.L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., Guisande, C., 2018. KnowBR: an application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. Ecol. Indic. 91, 241–248.

Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L., Werner, F., 2006. Techniques for cetacean-habitat modelling. Mar. Ecol. Prog. Ser. 310, 271–295.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Glob. Ecol. Biogeogr. 20, 789–802.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol. Lett. 9, 1136–1145.

Sánchez-Fernández, D., Lobo, J.M., Hernández-Manrique, O.L., 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. Divers. Distrib. 17, 163–171.

Shcheglovitova, M., Anderson, R.P., 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecol. Model. 269, 9–17.

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Tulloch, A.I.T., Sutcliffe, P., Naujokaitis-Lewis, I., Tingley, R., Brotons, L., Ferraz, K.M.P., Possingham, H., Guisan, A., Rhodes, J.R., 2016. Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. Biol. Conserv. 199, 157–171.

Valavanis, V.D., Pierce, G.J., Zuur, A.F., Palialexis, A., Saveliev, A., Katara, I., Wang, J.J., 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia 612, 5–20.

Van Sickle, J., 2010. Basic GIS Coordinates. CRC Press, Boca Raton, Florida.

Velasco, J.A., González-Salazar, C., 2019. Akaike information criterios should not be a "test" of geographical prediction accuracy in ecological niche modelling. Ecol. Inform. 51, 25–32.

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

Washington, W.M., Parkinson, C.L., 2005. An Introduction to Three-Dimensional Climate Modelling. Scientific American Library.

Webb, T.J., Vanden Berghe, E., O'Dor, R., 2010. Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep Pelagic Ocean. PLoS ONE 5, e10223.