



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS -
BIODIVERSIDADE NEOTROPICAL (PPGBIO)

Eduardo Motta Carelli Minsky

Leave out or put in - selecting input data to improve ecological
niche models applied to conservation and climate change
analysis: an approach using the Atlantic Goliath Grouper,
Epinephelus itajara (Perciformes)

Rio de Janeiro

2021

Leave out or put in - selecting input data to improve ecological niche models applied to conservation and climate change analysis: an approach using the Atlantic Goliath Grouper, *Epinephelus itajara* (Perciformes)

Eduardo Motta Carelli Minsky

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Biodiversidade Neotropical) da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para obtenção do título de Mestre em Ciências Biológicas.

Orientadora: Profa. Dra. Maria Lucia Lorini

Coorientador: Prof. Dr. Marcos de Souza Lima Figueiredo

Rio de Janeiro

2021

Catálogo informatizada pelo(a) autor(a)

M663	<p>Motta Carelli Minsky, Eduardo</p> <p>Leave out or put in - selecting input data to improve ecological niche models applied to conservation and climate change analysis: an approach using the Atlantic Goliath Grouper, <i>Epinephelus itajara</i> (Perciformes) / Eduardo Motta Carelli Minsky. -- Rio de Janeiro, 2021.</p> <p>171</p> <p>Orientador: Maria Lucia Lorini. Coorientador: Marcos de Souza Lima Figueiredo. Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Ciências Biológicas, 2021.</p> <p>1. Biogeografia. 2. Modelagem de Nicho Ecológico. I. Lorini, Maria Lucia, orient. II. de Souza Lima Figueiredo, Marcos, coorient. III. Título.</p>
------	--



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO – UNIRIO

Instituto de Biociências - IBIO

Programa de Pós-Graduação *Stricto Sensu* em Ciências Biológicas
(Biodiversidade Neotropical) - PPG BIO

ATA DA DEFESA DA DISSERTAÇÃO N°

Aos trinta dias do mês de junho de dois mil e vinte e um, às 09h00min, realizou-se na sala virtual do Google Meet, a defesa da dissertação intitulada "*Leave out or put in - selecting input data to improve ecological niche models applied to conservation and climate change analysis: an approach using the Atlantic Goliath Grouper, Epinephelus itajara (Perciformes)*" do discente Eduardo Motta Carelli Minsky do Mestrado em Ciências Biológicas da Universidade Federal do Estado do Rio de Janeiro (UNIRIO), como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências Biológicas (Biodiversidade Neotropical). A presente dissertação foi orientada pela Dra. Maria Lucia Lorini (Presidente da Banca) e coorientada pelo Dr. Marcos de Souza Lima Figueiredo. A Banca Examinadora foi constituída pelos seguintes membros, a saber: Dr. Bruno Moreira de Carvalho, e Dr. Luciano Neves dos Santos, com suplência da Dra. Tainá Rocha e do Dr. Rafael da Rocha Fortes. Após arguir o mestrando e considerar que o mesmo demonstrou capacidade no trato do tema escolhido e sistematização dos dados, a Banca Examinadora houve por bem aprová-lo com conceito A.

Rio de Janeiro, 30 de junho de 2021.

Dra. Maria Lucia Lorini
(Universidade Federal do Estado do Rio de Janeiro - UNIRIO)
Presidente da Banca

Dr. Bruno Moreira de Carvalho
(ISGlobal - Barcelona Institute for Global Health)

Dr. Luciano Neves dos Santos
(Universidade Federal do Estado do Rio de Janeiro - UNIRIO)

Eduardo Motta Carelli Minsky

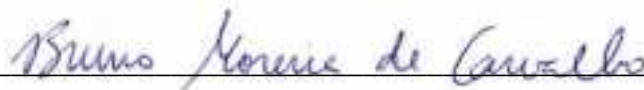
Leave out or put in - selecting input data to improve ecological niche models applied to conservation and climate change analysis: an approach using the Atlantic Goliath Grouper, *Epinephelus itajara* (Perciformes)

Dissertação apresentada ao curso de Mestrado em Ciências Biológicas do Programa de Pós-Graduação em Ciências Biológicas (Biodiversidade Neotropical) da Universidade Federal do Estado do Rio de Janeiro, no dia 30 de junho de 2021, como requisito para a obtenção do título de Mestre em Ciências Biológicas.

Banca examinadora:



Dra. Maria Lucia Lorini (Universidade Federal do Estado do Rio de Janeiro - UNIRIO) Presidente da Banca



Dr. Bruno Moreira de Carvalho (ISGlobal - Barcelona Institute for Global Health)



Dr. Luciano Neves dos Santos (Universidade Federal do Estado do Rio de Janeiro - UNIRIO)

Resumo

A oferta de pacotes de software prontos para uso e o aumento da disponibilidade de geoinformação digital sobre dados de ocorrência e dados ambientais impulsionaram consideravelmente a aplicação de modelos de nicho ecológico (MNEs) / modelos de distribuição de espécies (MDSs) nas últimas décadas, permitindo seu uso mais amplo para informar ações de conservação e gestão e para quantificar os impactos das mudanças globais. No entanto, embora os dados de entrada para MNE estejam cada vez mais disponíveis, definir procedimentos para minimizar as incertezas (espaciais, ambientais, temporais e taxonômicas) associadas aos dados de ocorrência e selecionar os preditores ambientais corretos continua sendo um desafio. Esta dissertação de mestrado, estruturada em três capítulos, concentra-se no desenvolvimento de abordagens para selecionar dados de entrada para ENMs com o objetivo de melhorar os modelos aplicados à análise de mudanças climáticas e conservação. Para ilustrar nossas abordagens, usamos um peixe ameaçado icônico, a Garoupa-Golias do Atlântico (*Epinephelus itajara*), como espécie modelo. No primeiro capítulo, desenvolvemos um arcabouço de várias etapas que foi capaz de apoiar a decisão sobre o uso de registros não confiáveis na modelagem de nicho ecológico para a espécie-alvo. Mostramos que os registros de ocorrência com incertezas diminuem o desempenho dos modelos, aumentando o erro de omissão e diminuindo sua capacidade de projetar os modelos do espaço ambiental para o espaço geográfico, levando a um baixo poder de predição de áreas adequadas. No segundo capítulo, comparamos modelos desenvolvidos com diferentes combinações de preditores ambientais para avaliar o efeito da adição de variáveis de habitat no desempenho de MNEs. Descobrimos que os preditores de habitat em combinação com o clima têm uma forte influência na precisão dos MNEs e nas previsões de adequabilidade. O modelo combinando clima / salinidade e preditores de habitat para diferentes habitats (EnvHabs) mostrou o melhor desempenho. Finalmente, no terceiro capítulo, aplicamos o melhor modelo (EnvHabs) identificado no segundo capítulo para avaliar os impactos das mudanças climáticas futuras na distribuição geográfica de *E. itajara* e discutir suas implicações do ponto de vista da conservação da espécie.

Palavras-chave: Biogeografia, Modelagem de Nicho Ecológico

Abstract

The offer of ready-to-use software packages and increasing availability of digital geoinformation about occurrence and environmental data have considerably boosted the application of ecological niche models (ENMs)/species distribution models (SDMs) in the past decades, greatly enabling their broader use for informing conservation and management, and for quantifying impacts from global change. However, although input data for ENM are increasingly available, to define procedures to both minimize uncertainties (spatial, environmental, temporal and taxonomic) associated to the occurrence data and to select the correct environmental predictors remain a challenge. This master thesis, structured into three chapters, focuses on developing approaches to select input data for ENMs aiming to improve models applied to climate change and conservation analysis. To illustrate our approaches, we used an iconic threatened fish, the Atlantic Goliath Grouper (*Epinephelus itajara*), as a model species. In the first chapter, we developed a multiple steps framework that was able to support decision about the use of unreliable records in ecological niche modeling for the target species. We showed that uncertain occurrence records decrease the models' performance, increasing their omission error and decreasing their ability to project the models from the environmental space to the geographical space, leading to low power to predict suitable areas. In the second chapter, we compared models developed with different combinations of environmental predictors to assess the effect of adding habitat variables on ENMs performance. We found that habitat predictors in combination with climate have a strong influence on ENMs accuracy and suitability predictions. The model combining climate/salinity and habitat predictors for different habitats (EnvHabs) showed the best performance. Finally, in the third chapter, we applied the EnvHabs best model (EnvHabs) identified in the second chapter to assess the future climate change impacts on the geographic distribution of *E. itajara* and discuss its implications from a conservation point of view.

Keywords: Biogeography, Ecological Niche Modeling

List of figures

Chapter 1: A framework to support decision about uncertainty occurrence records in ecological niche models

Figure 1.1: Workflow proposed for assess the effect of uncertainties and biases in occurrence data for ENMs. The workflow procedures are grouped into three main steps: (1) occurrence data filtering (blue), (2) Ecological niche modeling (green), and (3) model and datasets evaluation and comparison (orange).

Figure 1.2: Distribution of the occurrence records of *Epinephelus itajara*. Green dots - reliable records (Rd), red dots - unreliable records (Pd).

Figure 1.3: Variation in model performance statistics among occurrence datasets and methods.

Figure 1.4: Contribution of the environmental variables to model performance measured by AUC. Values are the mean and standard of all 250 models generated for each occurrence dataset (FullSet, Rd, Pd).

Figure 1.5: Response curves (black solid line) of the ensemble models generated for the three occurrence datasets (FullSet, RdPd, PdRd). The gray shadow represents the confidence interval considering all 250 replicates for each of three occurrence datasets.

Figure 1.6: Environmental suitability for *Epinephelus itajara* predicted by the ensemble models generated with the three occurrence datasets: FullSet (A), RdPd (B), and PdRd (C). Environmental suitability values (≥ 0.5) are represented from dark blue (minimum value) to red (maximum value).

Figure 1.7: Temporal coverage of the occurrence datasets of *Epinephelus itajara* showing the percentage of unreliable (Pd) and reliable (Rd) records along the years.

Figure 1.8: Environmental space of occurrence datasets of *Epinephelus itajara* summarized by the Principal Component Analysis composed by the five environmental variables used in the models. Red dots - unreliable records (Pd). Green dots - reliable records (Rd). Blue dots - environmental conditions for the area accessible to the

species, delimited by the ecoregions where the species is known to occur and adjacent, and sampled through 10,000 randomly distributed points. Dots are enclosed by ellipses encompassing 95% of the data.

Chapter 2: Adding habitat predictors to ecological niche models: an approach to improve model accuracy and suitability predictions

Figure 2.1: Spatial representation of the marine ecoregions (in light yellow) and the 92 presence records (green dots) selected for ecological niche modeling.

Figure 2.2: Euclidean distance variable for Seagrass habitat. Blueish colors indicate shorter distances to the appropriate seagrass areas and reddish colors indicate greater distances to the appropriate seagrass areas. The shortest distance found is 0 km and the largest distance is 4,973,768,000 km.

Figure 2.3: Euclidean distance variable for the Mangrove habitat. Blueish colors indicate shorter distances to the appropriate areas of seagrass and reddish colors indicate greater distances to the appropriate areas of mangrove. The shortest distance found is 0 km and the largest distance is 4,749,136,000 km.

Figure 2.4: Euclidean distance variable for Rockyreef habitat. Blueish colors indicate shorter distances to the appropriate areas of rockyreef and reddish colors indicate greater distances to the appropriate areas of rockyreef. The shortest distance found is 0 km and the largest distance is 3,644,155,000 km.

Figure 2.5: Euclidean distance variable for the Coralreef habitat. Blueish colors indicate shorter distances to the appropriate areas of coralreef and reddish colors indicate greater distances to the appropriate areas of coralreef. The shortest distance found is 0 km and the largest distance is 5,370,715,500 km.

Figure 2.6: Euclidean distance variable for the sum of all habitats. Blueish colors indicate shorter distances to the appropriate areas of all habitats and reddish colors indicate greater distances to the appropriate areas of all habitats. The shortest distance found is 0 km and the longest distance is 16,913,706,000 km.

Figure 2.7: Boyce Index P/E Ratio curve for the four models: (A) Env+SumHabs, (B) Env+Habs, (C) EnvOnly, (D) HabOnly. For all plots, the horizontal straight line is

where the P/E ratio is equal to 1. The dashed lines are the limits for the classification between High Suitability (HS) and Moderate Suitable (MS), i.e., habitat suitability values before the first dashed line are Not Suitable (NS), values between the line are considered MS and values after the second line are considered HS.

Figure 2.8: Environmental suitability for *Epinephelus itajara* predicted by the four ensemble models: (A) Env+SumHabs, (B) Env+Habs, (C) EnvOnly, (D) HabOnly. Dark blue colors represent High Suitability (HS) and light blue colors represent Moderate Suitable (MS).

Chapter 3: Applying ecological niche models optimized by habitat predictors to assess potential impacts of climate change on species distribution

Figure 3.1: Environmental suitability for *Epinephelus itajara* predicted by the ensemble model in the present. Dark blue = High Suitability (HS), Light blue = Moderate Suitability (MS).

Figure 3.2: Environmental suitability for *Epinephelus itajara* predicted by the ensemble model in the present. Dark blue = High Suitability (HS), Light blue = Moderate Suitability (MS).

Figure 3.2: Environmental suitability for *Epinephelus itajara* predicted by the ensemble model in the future. Dark blue = High Suitability (HS), Light blue = Moderate Suitability (MS).

Figure 3.3: Distribution of the stable areas for *Epinephelus itajara*. Green area represent stable areas for High Suitability (HS) pixels and yellow areas represent stable area for Moderate Suitability (MS) pixels.

List of Tables

Chapter 1: A framework to support decision about uncertainty occurrence records in ecological niche models

Table 1.1: Mean (\pm SD) values of the evaluation metrics for all the 750 models generated for *Epinephelus itajara* (250 replicates for each of three occurrence dataset).

Table 1.2: Two-factors factorial ANOVA results showing the effect of data partitioning (FullSet, RdPd, PdRd) and modeling methods (BRT, GLM, Maxent, RBF, SVM) on model performance metrics.

Table 1.3: Tukey Honest Significance Differences (HSD) post-hoc test results showing the effect of data partitioning (FullSet, RdPd, PdRd) on model performance metrics.

Table 1.4: Mean contribution (\pm standard deviation) of environmental variables to model construction of 250 replicates generated for each occurrence dataset (FullSet, Rd, Pd), measured by AUC. biogeo08 = mean annual salinity; biogeo12 = annual variance in salinity, biogeo14 = sea temperature of the coldest months; biogeo15 = sea temperature of the hottest months.

Table 1.5: Similarity, in the geographic space, of the modeling outputs generated by the three-occurrence dataset (FullSet, RdPd, PdRd).

Chapter 2: Adding habitat predictors to ecological niche models: an approach to improve model accuracy and suitability predictions

Table 2.1: Variables used to generate each ecological niche model. Salinity.Lt.min = Average minimum salinity records, Temperature Lt.max = Average maximum temperature records, Temperature.Lt.min = Average minimum temperature records, Dist.To.Mangrove = Euclidean distance variable for each presence cell predicted as suitable for the mangrove ENM, Dist.To.Rockyreef = Euclidean distance variable for each presence cell predicted as suitable for the Rockyreef ENM, Dist.To.Seagrass = Euclidean distance variable for each presence cell predicted as suitable for the Seagrass ENM, Dist.To.Any.Hab = variable that represents the distances to all habitats.

Table 2.2: Evaluation of the different ecological niche models for the Goliath grouper.

Table 2.3: Total of area predicted as suitable (S), highly suitable (HS) and moderately suitable (MS) for *Epinephelus itajara* and total area under Marine Protected Areas.

Table 2.4: Total predicted area within some pixel of human pressure.

Summary

Resumo	i
Abstract.....	ii
List of figures.....	iii
List of Tables	vi
General Introduction	1
Chapter 1: A framework to support decision about uncertainty occurrence records in ecological niche models.....	5
1.1 INTRODUCTION	5
1.2 MATERIALS AND METHODS.....	9
1.3.0 RESULTS	16
1.4.0 DISCUSSION	22
Chapter 2: Adding habitat predictors to ecological niche models: an approach to improve model accuracy and suitability predictions	31
INTRODUCTION	31
2.0 MATERIALS AND METHODS.....	35
3.0 RESULTS	43
4.0 DISCUSSION	48
Chapter 3: Ecological niche models predict, in a scenario of future climate change, a shift to high latitudes for a vulnerable species of grouper.....	54
3.1 INTRODUCTION	54
3.2 MATERIALS AND METHODS.....	59
3.3 RESULTS	60
3.4 DISCUSSION	64
General Conclusions	69
References.....	70
Appendices.....	102

General Introduction

Due to anthropogenic activities the Earth systems are rapidly undergoing changes of enormous magnitude, which already achieve geological significance (Lewis et al., 2015; Araújo et al., 2019). Humanity is causing a rapid loss of biodiversity and, consequently, the Earth's ability to support complex life (Bradshaw et al., 2021). The increasing aware that biodiversity changes have important effects on human welfare and ecosystem services (Díaz et al., 2019; Bradshaw et al., 2021) also increases the demand for our ability to forecast changes in biodiversity (Araújo et al., 2019). Forecasting the responses of biodiversity to multiple drivers of change is a difficult task, that often rely on the use of models to help risk assessment and support decision about conservation and management actions (Guisan et al., 2013; Araújo et al., 2019).

Among biodiversity modeling approaches applied to assess the impact of anthropogenic stressors and to hind/forecasting, one of the most used are the ecological niche models (ENMs) / species distribution models (SDMs) (Araújo et al., 2019). Discussions about the differences between the two terms can be found in Peterson and Soberón (2012) and Soberón et al. (2017). ENMs are correlative models that relate the species occurrence to environmental variables to define the species ecological niche in the environmental space, and then project it onto the geographical space to identify the potential distribution of the suitable environmental space for the species (Peterson et al., 2011; Bellard et al., 2012). It is then possible to project this niche for different geographical space, or for different timeframe in past or future climate scenarios, to determine the potential distribution of the suitable environmental space of the species (Peterson et al., 2011; Bellard et al., 2012).

Thanks to the offer of ready-to-use software packages and increasing availability of digital geoinformation related to environmental and species occurrence data, in the last decades ENMs have been widely applied to a broad set of conservation, ecological and evolutionary questions, including: discovery of new populations of known species, discovery of previously unknown species, spatial conservation prioritization, assessment of potential geographic ranges of invasive species, mapping risk of disease transmission, forecasting the effects of climate change on species distributions and on phylogenetic diversity, and identifying historical refugia for biodiversity (Peterson et al. 2011; Lorini et al., 2015; Villero et al., 2017; Araújo et al., 2019; Melo-Merino et al.,

2020; Zurell et al., 2020). The beginning and the most of the development of ENMs is broadly associated to terrestrial environments (Lorini et al., 2015; Villero et al., 2017; Araújo et al., 2019). Although the application of ENMs to aquatic species has been less frequent, in recent years this type of modeling has shown a growing trend in the marine realm (Robinson et al., 2017; Melo-Merino et al., 2020).

The processes of modelling species geographic distribution through ENMs rely on ecological niche theory (Soberón, 2007; Soberón and Nakamura, 2009; Peterson et al., 2011). The first niche concept adopted by the scientific community were proposed by Grinnell (1924), who took into account the environmental conditions and defined the ecological niche of a species as the suite of ecological conditions within which a species is capable of surviving and reproducing without immigrational subsidy. Next, Elton (1927) proposed that the interactions between species were crucial factors in the delimitation of the species niche, redefining the term ecological niche to refer to the functional role that a species plays in a community. Later, Hutchinson (1957) proposed a concept encompasses these two views and defines an ecological niche as a hypervolume, n-dimensional, whose axes encompass conditions, resources and interactions in which individuals of a species are able to survive, grow and reproduce. Hutchinson (1957) also proposed a distinction between the fundamental niche, the set of abiotic environmental conditions under which a species is able to persist indefinitely, and the realized niche, the part of the fundamental niche that is constrained by interactions with other species.

Following these ideas, species' ranges can be understood as resulting from three sets of factors (Soberón, 2007; Soberón and Nakamura, 2009; Peterson et al., 2011): (i) presence of environmental (abiotic or scenopoetic) conditions under which the species can establish, survive and reproduce; (ii) the biotic environment determined by the species interactions (e.g. resources, competition, predation, pathogens) in which species can persist, and (iii) the area that is accessible to the species via its movement or dispersal capabilities. The representation of these range determinants has been captured in the so-called BAM diagram (B = Biotic, A = Abiotic, M = Movement), which has become a conceptual model to design ENMs (Soberón, 2007; Soberón & Nakamura, 2009; Peterson et al., 2011, Barve et al., 2011). Therefore, species geographic ranges are the result of the dynamic interactions of those three sets of determinants. Due to the complexity surrounding biotic interactions (B) and because biotic interactions are hard to measure and represent in a spatially explicit format required by ENMs (Soberón and

Nakamura, 2009), most correlative ENMs have been restricted to representations of abiotic conditions (A). This approach is supported by the Eltonian Noise Hypothesis (Soberón and Nakamura, 2009; Peterson et al., 2011), which proposes that biotic interactions seldom constitute a significant constraint on the distributional potential of species on large geographic extents and coarse resolutions, like those generally used in ENMs. Indeed, ecological interactions may not play a dominant role at the coarse resolution ($>1\text{km}^2$) typically used in ENMs (Soberón, 2007; Hortal et al., 2010), and thousands of studies have demonstrated that ENMs can reach a good performance in describing the distribution of many species on the basis of A (but see Araújo et al., 2014).

Technically, the ENM process is supported by three fundamental pillars: (1) information about the species (physiological tolerance from occurrence data), (2) environmental variables (predictor variables) and (3) the analytical methods themselves (functions or models that relate species information to environmental predictors) (Lima-Ribeiro and Diniz-Filho, 2013). The procedures related to processing, filtering and selection of the occurrence data and predictor variables have great influence on the ENM performances and outputs (Araújo et al., 2019). Although occurrence data are increasingly available, unfortunately, not all data are of the same quality as they are subject to uneven sampling, taxonomic misidentification, errors in spatial coordinates and other data entry errors. Hence there is a need for adequate quality control aiming to minimize spatial, environmental, temporal and taxonomic uncertainties associated to the occurrence data (Araújo et al., 2019; Zizka et al., 2019; Zurell et al., 2020). Similarly, while environmental data are increasingly available, the selection of the environmental predictors is not always obvious. Indeed, although the choice of the correct predictors used for ENMs is crucial in determining suitable habitat for a species and hence their predicted distribution, this process is a persistent challenge (Petitpierre et al., 2017; Bosch et al., 2018; Araújo et al., 2019; Zurell et al., 2020).

This master thesis focuses on developing approaches to select input data for ENMs aiming to improve models applied to climate change and conservation analysis. To illustrate our approaches, we used the Atlantic Goliath Grouper (*Epinephelus itajara*) as the study system. This iconic threatened fish is a species of conservation concern, categorized as globally Vulnerable to extinction and very popular for fishing (Beroncini et al. 2018). The occurrence records for this species are associated with

different uncertainty levels. This tropical species presents dependence on different types of coastal habitats (estuaries, rocky shores, coral reefs). Those characteristics make *Epinephelus itajara* a particularly valuable model species for our proposals. The thesis is structured in three chapters that focused on two of the fundamental pillars of ENM, the occurrence data (Chapter 1) and predictor variables (Chapters 2 and 3). In the first chapter, we addressed the problem of deal with biases and different levels of uncertainties in occurrence datasets, and proposed a framework to support decision about to discard or use unreliable records in occurrences datasets. In the second chapter, we developed models with different combinations of environmental predictors to assess the effect of adding habitat variables on the performance of ENMs. At last, in the third chapter we applied the best model approach developed in the second chapter to improve the analysis of future climate change impacts on the range of *Epinephelus itajara*.

Chapter 1: A framework to support decision about uncertainty occurrence records in ecological niche models

1.1 INTRODUCTION

Information on species distribution is essential to answer biogeographic (Lomolino, 2004) ecological (Brown et al., 1996) and evolutionary (Holt, 2003) questions, and to support decisions regarding biodiversity management and conservation (Boitani et al., 2011; Jetz et al., 2012; Villero et al., 2017). In the last decades, the interest in understanding the distribution of global biodiversity has increased due to the need to protect it, in face to the continuous and widespread decline of biodiversity, and significant reduction of nature's capacity to contribute to people's well-being (MEA, 2005; IPBES, 2019; Halpern et al., 2019). Concomitantly, Ecological Niche Models (ENMs) have emerged as an interesting and efficient approach to analyze species geographic distributions, and have been used in a myriad of different applications (Guisan & Thuiller, 2005; Elith & Leathwick, 2009; Peterson et al., 2011; Villero et al., 2017). These correlative models relate the species occurrence to environmental variables to define the species ecological niche in the environmental space, and then project it onto the geographical space to identify the potential distribution of the suitable environmental space for the species (Peterson et al., 2011; Bellard et al., 2012).

The widespread application of ENMs is, in great part, driven by an increasing of digital freely accessible occurrence data (Anderson, 2012). A huge number of occurrence records has become available in biodiversity or citizen science databases, offering a great possibility for the application of ENM approach to many species (Guralnick et al., 2007; Varela et al., 2014). Occurrence records available in repositories such as the Global Biodiversity Information Facility (GBIF) are crucial for conservation, biogeography and macroecology studies as they are a permanent record of a species in a particular location at a particular point in time (Funk and Richardson, 2002). But data contained in these databases are a highly heterogeneous compilation of records. Such datasets include information from preserved specimens in museums and herbaria; university databases; governmental organization databases; NGO databases;

researchers field work; amateurs field work, resulting in a compilation of records gathered opportunistically or from hundreds of different surveys, each one designed with a different goal (Varela et al., 2014). Consequently, these datasets are prone to errors and biases (Graham et al., 2004; Ensing et al., 2013; Costa et al., 2015; Troudet et al., 2017).

The uncertainties of occurrence datasets are spatial, environmental, temporal and taxonomic (Boakes et al., 2010; Faith and Walker, 1996; Funk et al., 1999; Soberón et al., 2000; Sousa-Baena et al., 2014). Reliability of occurrence records, which primarily refers to the accuracy of the species' identification, is an obvious issue to ENM (Graham et al., 2004; McKelvey et al., 2008; Lozier et al., 2012; Ensing et al., 2013; Costa et al., 2015; Aubry et al., 2017). In addition, the dynamic nature of the accepted taxonomies can also contribute to misidentifications, as different names may be given to the same species depending on the date of the collect or taxonomic determination of the record (Graham et al., 2004; Newbold, 2010). Temporal mismatches of occurrences records and environmental predictors can impact significantly the results of ENMs (Roubicek et al., 2010). Positional uncertainties of the occurrence data can affect the performance and interpretation of species distribution models (Wieczorek et al., 2004; Graham et al., 2008; Osborne & Leitão, 2009; Soutan & Safi, 2017). However, all these uncertainties have received less attention than occurrence data biases (Gueta & Carmel, 2016).

Several studies show that there is uneven spatial coverage in sampling (Austin and Meyers, 1996; Cheng et al., 2013; Dennis and Thomas, 2000; Kadmon et al., 2009; Meyer et al., 2016, 2015; Reddy and Dávalos, 2003; Soberón et al., 2000) and there are evidences that these limitations may be driven mostly by the physical accessibility (e.g. Kadmon et al., 2004; Amano and Sutherland, 2013; Ballesteros-Mejia et al., 2013; Lin et al., 2015), but also by the proximity of researchers (Moerman & Estabrook, 2006); socio-economic differences (Newbold, 2010) and lack of financial resources (Ahrends et al., 2011; Amano and Sutherland, 2013; Vollmar et al., 2010), low international scientific cooperation (Amano and Sutherland, 2013); lack of access to a region for security reasons (Amano and Sutherland, 2013; Ballesteros-Mejia et al., 2013; Moerman and Estabrook, 2006) and finally focus on regions with a high level of endemic species or protected areas (Boakes et al., 2010; Yang et al., 2013). Sample bias in geographic space may not be a problem until the records cause an environmental bias (Austin and Meyers, 1996; Kadmon et al., 2009). However, bias in geographical space

has already been proven to cause bias in environmental space (e.g. Gonzalez, 2010; Funk & Richardson, 2002) and should be taken with extreme caution (Guisan and Thuiller, 2005). Bias in the environmental space in which the species' niche are modeled can cause over-representation of environmental conditions associated with regions of clustered sampling (Boria et al., 2014), misinterpretation of significance of environmental predictors (Kühn, 2007), and inflated estimates of model performance (Hijmans, 2012; Hijmans & Hall, 2016; Veloz, 2009). Geographic sample biases can lead to inaccurate inferences and predicted distributions of ENMs (Stolar & Nielsen, 2015), and several approaches exist to account for it (e.g. Phillips et al., 2009; Chakraborty et al., 2011; Fithian & Hastie, 2013; Renner & Warton, 2013; Varela et al., 2014, Warren et al., 2014, Boria et al., 2014, Fourcade et al., 2014, Fithian et al. 2015, Stolar & Nielsen, 2015; Aiello-Lammens et al. 2015), but none of these approaches solved the problem completely.

There is a general agreement in the literature that uncertainties and biases in occurrence data have negative effects on the ENMs and that the modelers should deal with them. On the other hand, there is also a great consensus that low sample size decreases accuracy of ENMs. Model performance is known to rapidly decrease for sample sizes smaller than 20, although the minimum number of species occurrences needed to produce accurate models still remains debatable (Stockwell & Peterson, 2002; Drake et al., 2006; Hernandez et al., 2006; Pearson et al., 2007; Papeş & Gaubert et al., 2007; Loiselle et al., 2008; Wisz et al., 2008; Mateo et al., 2010; Marini et al., 2010; van Proosdij et al., 2016). Consequently, there is a trade-off between attaining a minimum level of uncertainties/biases and maintaining the maximum number of occurrences in the dataset. This is challenge for species whose distributions present disjunctions or high environmental heterogeneity, which are prone to environmental biases, and for rare, poorly known or elusive species, which are prone to uncertainties and small number of occurrences. Implications of this decision about how to treat uncertainties and biases of occurrence data can be especially severe for threatened species, when the ENM predictions may be used to guide management and conservation actions (Sofaer et al., 2019). Moreover, modelers working with tropical species often have to face the problem of small number of available occurrences (Cayuela et al., 2009; Feeley & Silman, 2011; Feeley et al., 2015). At the same time, it is precisely in tropical regions that information on the distribution of biodiversity is most needed (Funk & Richardson,

2002; Cayuela et al., 2009; Feeley & Silman, 2011; Feeley et al., 2015), a lack of knowledge referred as the “data void” problem by Feeley and Silman (2011).

Although several approaches regarding how to deal with occurrences geographic biases are available in the literature, there is a lack of approaches to guide the modelers how to deal with distinct types of uncertainties and biases together. Moreover, unlike to well-studied effects of the geographic biases, much less is known about the impact of temporal biases. One of those few studies (Reside et al., 2011) showed that including historical and low-resolution records would decrease the accuracy and quality of ENMs and that deleting these records would be the best option. But what happens when part of the geographic distribution information is based on old, single source, low-resolution occurrence records? This kind of Wallacean shortfall is presented by many species, especially in tropical regions (Cayuela et al., 2009; Feeley & Silman, 2011; Feeley et al., 2015), rising the question about what do with these low reliability records. Would the inclusion of these unreliable records in a model to improve sample size, and environmental and geographic coverage compensate the increased chance of error caused by coarse or inaccurate data? Or should we exclude those records before starting the modeling process?

Here, we addressed this problem analyzing the occurrence data for the Atlantic Goliath Grouper (*Epinephelus itajara*). The species is exemplar to illustrate such a kind of Wallacean shortfall due to its amphi-Atlantic distribution, with disjunct populations in Western and Eastern Atlantic coasts, and uncertainties associated to African occurrence records (Craig, 2015). The historical distribution of this species encompasses the Western Atlantic Ocean coast from Brazil to United States, and Eastern Atlantic coast between Senegal and Angola. *Epinephelus itajara* is the largest grouper in the Atlantic (up to 2 m and 400 kg), being currently assessed as globally Vulnerable to extinction according to the IUCN Red List of Threatened Species (Bertoncini et al., 2018). The Atlantic Goliath Group is suspected to have declined on a global-level and might have been locally extinct in the African continent due to overfishing (Craig, 2015). Furthermore, reliable and recent records are geographically associated to the Western Atlantic whilst poor data are concentrated in the Eastern Atlantic. Sampling and conservation efforts are higher in the western portion of its distribution compared to the eastern portion (Craig et al., 2009) and available occurrence data in Eastern Atlantic are practically restricted to online databases (GBIF).

Our study workflow encompasses procedures in multiple steps to deal with biases and with different levels of uncertainties in occurrence datasets of *Epinephelus itajara*. We applied a multifaceted approach to evaluate the models (different performance metrics, spatial prediction ability, importance of predictors, behavior of response curves), and the occurrence datasets (temporal and environmental coverage) to support decision about to discard or use unreliable records in ecological niche / species distribution modeling. Within this framework, we assessed the uncertainty and biases of the occurrence records and tested two hypotheses:

(H1) Uncertainty level in occurrence datasets will affect ENMs performance, even after filtered to eliminate records flagged as very high uncertainty and reduce bias and spatial autocorrelation.

(H2) Model accuracy will be lower in ENMs calibrated with occurrence datasets with high uncertainty level.

1.2 MATERIALS AND METHODS

To deal with and assess the effect of uncertainties and biases in occurrence data of *Epinephelus itajara* in ENMs, we divided the analysis workflow into three main steps (Figure 1.1): (1) we classified the uncertainty level of the occurrence data, applied filter to remove very high uncertainty records and spatial biases, and created the occurrence datasets with different levels of reliability to be used in modeling; (2) generated ENMs to each occurrence dataset; and (3) evaluated models accuracy and occurrence datasets through multiple aspects, then make decision about unreliable records (Figure 1.1).

1.2.1 Occurrence data filtering

We compiled occurrence data for *E. itajara* from online databases, literature and specialists' personal communications. Online repositories included: Global Biodiversity Information Facility (<http://www.gbif.org>); Ocean Biogeographic Information System (<https://obis.org>); FishBase (<https://www.fishbase.de>); SpeciesLink (<http://splink.cria.org.br>); Sistema de Informação sobre a Biodiversidade Brasileira

(<http://www.sibbr.gov.br>). The literature survey for published papers was conducted online through Web of Science and Google Scholar search engines using "*Epinephelus itajara*" as search terms, resulting in five principal sources (Artero et al., 2015; Creed et al., 2007; Damasceno et al., 2015; Tzadik et al., 2017; Zapelini et al., 2017). We contacted more than 20 fish and grouper experts and occurrence data were provided by six from these specialists. The compiled occurrence dataset totaled 18,424 records.

We evaluated the compiled data for taxonomic, spatial and temporal uncertainties and classified them into four levels of uncertainty: low, medium, high and very high. Records misidentified as *E. itajara* (Craig et al., 2009), with geographic inconsistencies (records located on land; with incompatible latitude and longitude; with coordinates inconsistent with associated locality; or too far from the species' Extent of Occurrence polygon provided by IUCN) or old records (dated from before 1960) were assigned to the very high uncertainty level. Records from literature or from localities provided by specialists were assigned to the low uncertainty level. Records from repositories were classified according to their proximity to the records from literature or provided by specialists: records with higher proximity (<50km) were assigned to the medium uncertainty level, while records with lesser proximity (≥ 50 km) were assigned to the high uncertainty level. We then discarded the records assigned to the very high uncertainty level and after this procedure the occurrence dataset totaled 8,453 records.

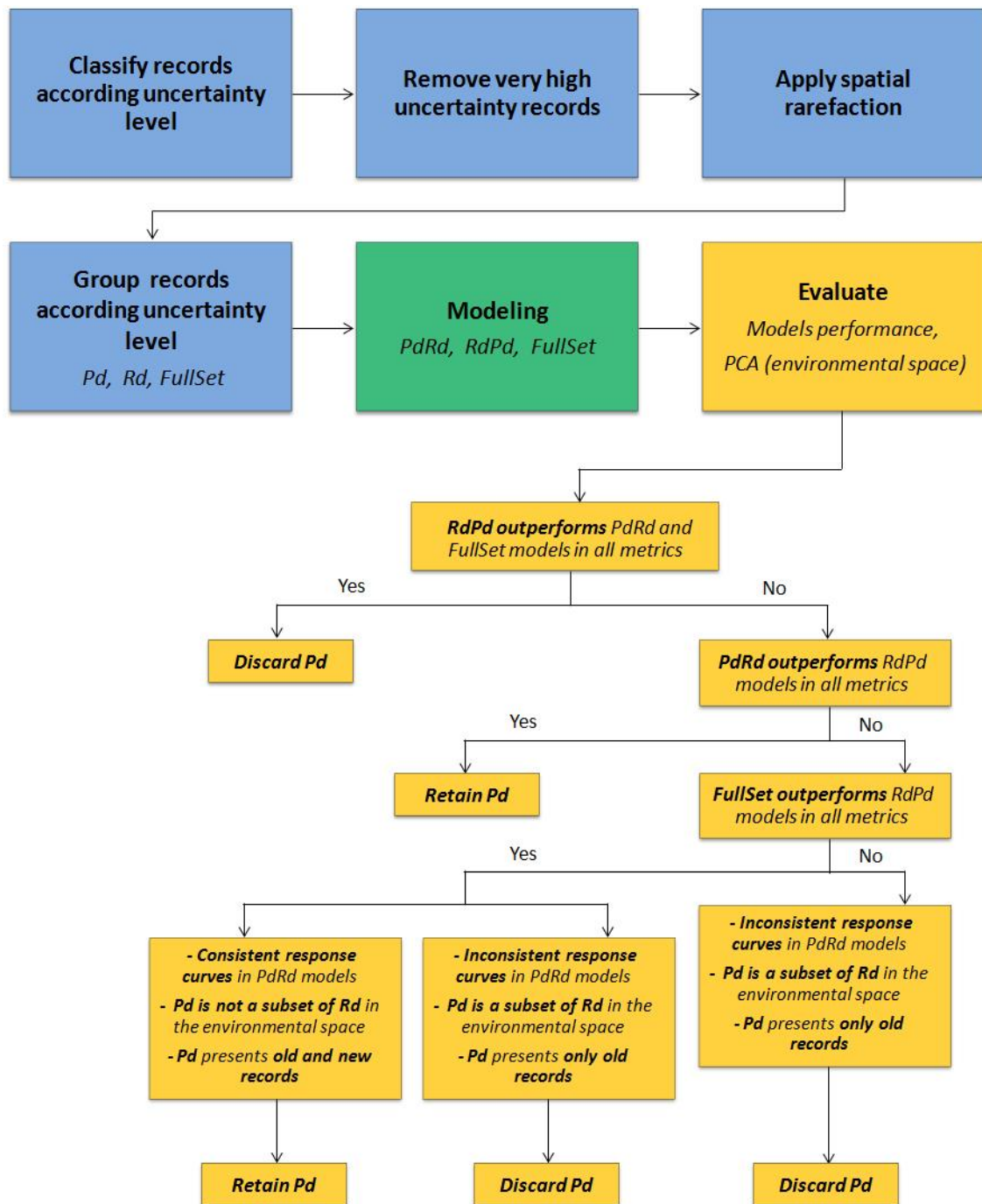


Figure 1.1: Workflow proposed for assess the effect of uncertainties and biases in occurrence data for ENMs. The workflow procedures are grouped into three main steps: (1) occurrence data filtering (blue), (2) Ecological niche modeling (green), and (3) model and datasets evaluation and comparison (orange).

Spatial autocorrelation (SAC) and sample bias can lead to niche model uncertainties (F. Dormann et al., 2007). Specifically, SAC and sample bias violate the assumption that records are independent and randomly distributed, which can lead to model inflation, thus, raising Type I Error. Many methods have been tested to mitigate

such problems, but there is no consensus about their efficiency (El-Gabbas and Dormann, 2018a, 2018b; F. Dormann et al., 2007). Here we applied a spatial rarefaction that eliminates records closer to each other than the expected mean distance in random sets with the same number of points, implemented by the SDMToolbox tool (Brown, 2014). We used Average Nearest Neighbor analysis (Chen & Getis, 1998) to estimate the expected mean distance for a hypothetical random distribution of a set with 8,453 points (= 50.4 km). Therefore, we defined 50 km as a distance threshold for rarefaction in the *E. itajara* dataset. We also used Average Nearest Neighbor Ratio (expected/observed mean distance) to test the presence of clustering (ANNR < 1) or dispersion (ANNR > 1) patterns in datasets before and after the spatial rarefaction. To avoid distortions, distance calculations were computed with data in equidistant projection. All procedures were performed in ArcMap 10.5 (ESRI 2018).

Subsequently, we created three datasets based on records uncertainty level: Reliable Data (Rd) - grouping records assigned as the low or medium uncertainty level; Unreliable Poor Data (Pd) - grouping records assigned as the high uncertainty level; and Full Data (FullSet) - grouping all records.

1.2.2 Ecological niche modeling

We used the correlative ENM approach to associate occurrence data for *E. itajara* to environmental predictors and then identify areas that are ecologically suitable for the species. We ran ecological niche models for each occurrence dataset (Rd, Pd, FullSet) developed in the previous step.

Considering that the species occurs in shallow and coastal waters, we delimited a calibration area for ENMs based on the World Marine Ecoregions (SPALDING et al., 2007) in which species was present, plus those ecoregions immediately adjacent (see Appendix 1, Figure S1, Table S1). This calibration area has a larger spatial extent than the species distribution range, but restricts modeling to accessible area over time, because some model performance evaluation metrics tend to increase artificially as the geographical extent of the analysis also increases (Barve et al., 2011). We used the limits of the selected ecoregions to clip the environmental predictors.

Environmental predictors at a resolution of 5 arc-minutes (~10km) were obtained from the MARSPEC database (Sbrocco and Barber, 2013). We have compiled

an initial set of 11 predictors, including 10 climate variables representing annual averages, extremes and annual variances for ocean surface temperature (SST) and surface salinity (SSS), plus one geophysical variable representing bathymetry. These predictors were selected to represent a spectrum of environmental characteristics that can be related to ecological features of the target species and are often used in modeling distribution of marine species (Sandman et al., 2013; Bradie & Leung, 2017). The final selection of variables was based on three main aspects: biological relevance to influence marine fish (Sandman et al., 2013; Bradie & Leung, 2017; Goodyear et al., 2017) and *E. itajara* distribution (Gilmore et al., 1978; Eklund & Schull, 2001; Frias-Torres, 2006; Koenig et al. 2007; Chapman et al. 2014; Shideler et al., 2015); reduction of collinearity of the variables of the original set; and greater contribution to the model. Collinearity among predictors can decrease the efficiency and also increase the uncertainty of niche models (De Marco and Nóbrega, 2018). Reduction of collinearity was evaluated through a pairwise Spearman's correlation matrix and variables most correlated ($r^2 \geq |0.7|$) were excluded from the final predictors set. Variables' contribution was assessed through a Jackknife approach using the models generate with all variables in a preliminary phase (see details in Appendix 2, Table S2). The three-criteria selection resulted in a final set of five environmental predictors to be used in ENMs: bathymetry of the ocean floor; bioge08: mean annual salinity; bioge012: annual variance in salinity, bioge014: sea temperature of the coldest months; bioge015: sea temperature of the hottest months.

ENMs were generated using several algorithms combined in a final ensemble model to reduce uncertainties (Araújo & New, 2007; Qiao et al., 2015). A total of five algorithms were chosen to build the models: (i) Generalized Linear Model (GLM), (ii) Boosted Regression Tree (BRT), (iii) Maximum Entropy (Maxent), (vi) Support Vector Machine (SVM) and (v) Radial Basis Function (RBF). Together these algorithms cover a considerable range of different widely used modeling techniques and are among the best performing (Babak Naimi unpublished data). Modeling procedures were conducted in the "sdm" package (Naimi and Araújo, 2016) on the R platform (R Core Team, 2016). We created three sets of models: (i) RdPd - models calibrated with the low and medium uncertainty records (Rd) and independent validated with the high uncertainty (Pd) records, (ii) PdRd - models calibrated with high uncertainty records (Pd) and independent validated with low and medium uncertainty (Rd) records and (iii) FullSet - calibration and validation with full dataset partitions (FullSet). For each algorithm we

ran fifty replicas of models created from the bootstrapping method. Except Maxent, the other four algorithms use presence and absence or pseudo-absence to fit models. Considering that the number of pseudo-absences can influence the accuracy (ability to predict suitable or unsuitable areas) of the models (Barbet-Massin et al., 2012), we performed a sensitivity analysis to identify the better option of pseudo-absence-to-presence ratio. We used a 1:1 ratio to generate pseudo-absences, following the results of the preliminary sensitivity analysis (see details in Appendix 3, Table S3). Then we create pseudo-absences, which were randomly generated in the geographic space of calibration, excluding cells with detected presence plus a 50 km buffer. For Maxent we generated a set of 10,000 randomly distributed background points.

1.2.3 Models and datasets evaluation and comparison

We evaluated the performance of the 750 models generated (5 algorithms * 50 replicates * 3 occurrence datasets) using a combination of four different one-threshold-independent and -dependent evaluation metrics. From these discrimination metrics, the Area Under the Curve (AUC) of the Receiver-Operating Characteristic (Fielding & Bell, 1997); is one-threshold-independent. An AUC score of 1 indicates a perfect fit of the data, 0.5 is no better than random (Elith et al., 2006). Model performance and AUC values can be related as follows: poor to fair (0.5 – 0.7), moderate (> 0.7 – 0.8), good (> 0.8 – 0.9) and very good (> 0.9) (Swets, 1988; Franklin, 2010) The other three metrics, Sensitivity, Omission Error and True Skill Statistics (TSS) (Allouche et al., 2006), are one-threshold-dependent and were generated in the confusion matrix using the threshold that maximize TSS (MAX TSS) (Liu et al., 2013). This threshold is one the most recommended method and has been proved valid to use when pseudo-absences are used instead of true absences (Liu et al., 2013). Sensitivity is the detected presence ratio predicted as such, so it quantifies omission error (Type II Error). The TSS is calculated by the formula Sensitivity + Specificity - 1, i.e., it takes into account errors of omission and commission, ranges from - 1 to + 1, where + 1 indicates a perfect classification model, zero or less indicate performance no better than random (Allouche et al., 2006). At a given threshold, a model performs accurately if it scores a TSS > 0.5 (Allouche et al., 2006; Liu et al., 2011). Model performance and TSS values can be related as follows: poor (< 0.4), moderate (0.4 > 0.6), good (0.6 - 0.75), and very good (> 0.75)

(Landis & Koch, 1977; Capinha et al., 2014). We did not take Specificity into account, as we were interested in assessing whether our models would be able to detect occurrence records as such (true positive rate). Therefore, to our aim the more important evaluation metrics to accounting for were Sensitivity and Omission Error (Type II Error). We considered plausible models those reached TSS values of 0.5 or higher (Silva et al., 2014). Next, we built consensus ENMs using only plausible models (TSS \geq 0.5), and applying TSS weighted average ensemble method (Araújo & New, 2007; Marmion et al., 2009).

To evaluate the uncertainty in evaluation metric generated by modeling components (data partitioning vs. modeling methods) - i.e., if each evaluation metric significantly differed among models generated for each of the three occurrence datasets (RdPd, PdRd and FullSet) along the five modeling methods (BRT, GLM, Maxent, RBF, SVM) - we used a two-factor factorial ANOVA (Sensitivity / Type II Error / TSS / AUC ~ methods * data group) and Tukey Honest Significance Differences (HSD) *post-hoc* test (Fournier et al., 2017).

Comparisons of the models in the environmental space were carried out by analyzing the response curves and the variable importance associated to RdPd, PdRd and FullSet models. We evaluated the relative importance of each variable in predicting the distribution of the species in each model using the "getVarImp" function in the "sdm" package. The response curves of important predictors, which describe variation of the species suitability along the gradient of each variable, were also plotted to identify inconsistent behavior, such as erratic curves or large standard deviations and confidence intervals (Hannemann et al., 2016).

To assess the extent of change in predicted suitability (in the geographical space) caused by differences among occurrence datasets, we used Schoener's D statistic. We computed pairwise Schoener's D for the raster predictions of the three models (RdPd, PdRd, FullSet) in R package 'dismo' (Hijmans et al., 2016). Schoener's D ranges from 0 (no overlap in predicted suitability) to 1 (complete overlap, identical predictions) and provides a measure of the similarity of two modeling outputs in the geographic space (Warren et al., 2008). We also compared the ability of consensus models to predict suitable areas for *E. itajara* within the species' EOO under a fixed threshold (suitability value \geq 0.5). To avoid distortions, area calculations were performed with data in the World Miller Cylindrical equivalent projection. We also evaluated occurrence data sets regarding temporal and environmental coverage. To analyze the temporal coverage we

computed the ratio between time span of the records in the dataset and time span of all records. To analyze the environmental coverage, we overlapped the reliable (Rd) and unreliable (Pd) occurrence records with the environmental predictors used in the modeling, and extracted the value for each pixel. Then we performed a Principal Component Analysis to associate the occurrence records to the environmental predictors. PCA was used to identify the internal structure of the data and to investigate the similarity between reliable (Rd) and unreliable (Pd) occurrence datasets in the environmental space. This technique is used to represent the "environmental niche" occupied by the species (Janžekovič & Novak, 2012).

Finally, we developed guidelines to the decision about discard or retain unreliable records (Pd). Following the workflow, if the analyses in the evaluation step indicate that: (i) models fitted with data filtered to reduce uncertainty and biases (RdPd) outperformed those fitted with unfiltered data (PdRd), especially in Sensitivity, Omission Error; (ii) PdRd models presented inconsistent response curves; (iii) Pd represented a subset of Rd in the environmental space, and (iv) Pd presented low temporal coverage with only old records, then we propose discard the Pd records and use the RdPd model results. Otherwise, if the evaluation analysis indicates that: (v) models fitted with filtered data (RdPd) did not outperform those fitted with unfiltered data (PdRd), especially in Sensitivity, Omission Error; (vi) PdRd models presented consistent response curves; (vii) Pd did not represent a subset of Rd in the environmental space, and (viii) Pd presented a reasonable temporal coverage with old and new records, then we propose maintain the Pd records and use the FullSet model results.

1.3.0 Results

1.3.1 Occurrence filtered datasets

From the total geo-referenced unique occurrences records of *Epinephelus itajara* compiled from all sources (n = 14,027), more than one third represented records assigned to very high uncertainty level (n = 5,574). These very unreliable records were discarded and this procedure filtered out 39.74% of the compiled data. The remained

dataset was reduced to 8,453 occurrences, where those assigned to low, medium and high uncertainty level summed 65, 8319, and 69 records, respectively.

Subsequently, the spatial rarefaction procedure applied for this filtered *E. itajara* dataset (distance threshold = 50 km, based on a expected mean distance = 50.4 km, estimated through an average nearest neighbor analysis) was able to remove the clustering pattern presented by these 8,453 occurrences. Before spatial rarefaction the occurrence dataset presented a significantly clustered pattern (Nearest Neighbor Ratio: 0.53, z-score: -18.23, p-value: < 0.0001). After rarefaction, the occurrence dataset (138 records retained) did not differ from a random pattern (Nearest Neighbor Ratio: 1.02, z-score: 0.35, p-value: 0.730). Cumulative with the first filter, rarefaction procedure filtered out 99.02% of the compiled data.

At the end, three datasets were created with the 138 filtered records (Figure 1.2): Reliable Data (Rd) - grouping the 83 records assigned as the low or medium uncertainty level; Unreliable Poor Data (Pd) - grouping the 55 records assigned as the high uncertainty level; and Full Data (FullSet) - grouping all records. All the datasets presented a sample size at least twofold higher than the minimum size considered prone to decrease model performance. Rd and Pd presented a geographic disjunct distribution, in Western and Eastern Atlantic, respectively (Figure 1.2).

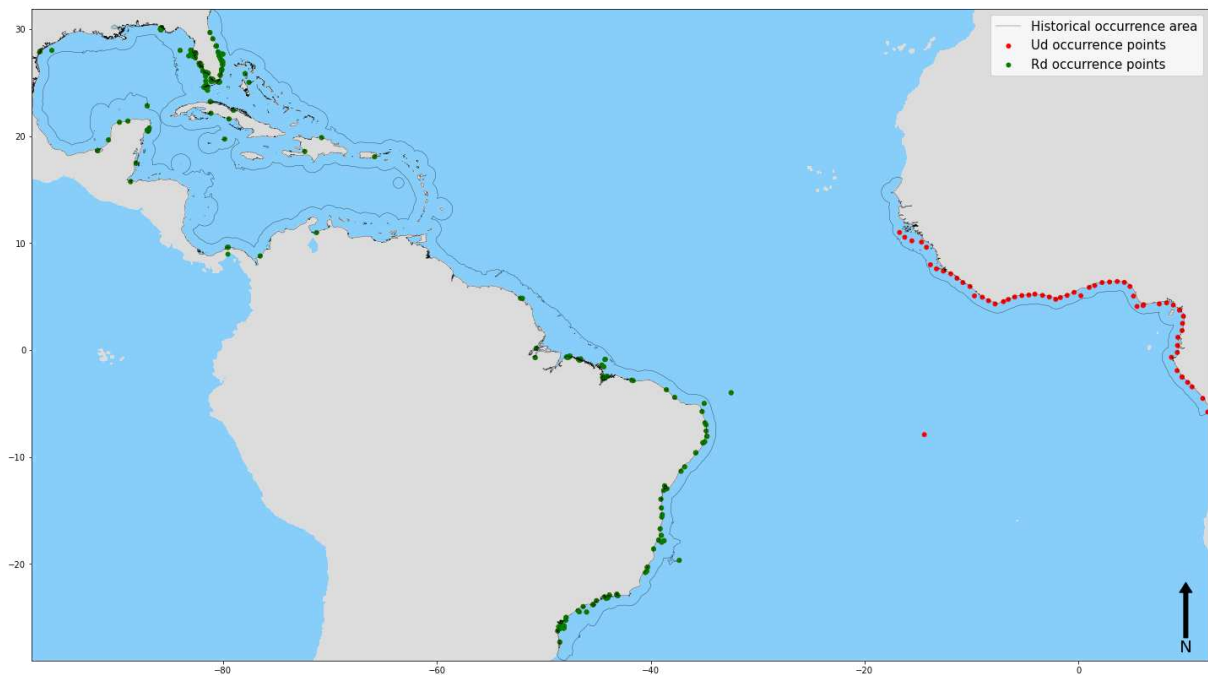


Figure 1.2: Distribution of the occurrence records of *Epinephelus itajara*. Green dots represent data collected from online databases (<50km proximity from reliable occurrence records), from literature and from localities provided by specialists. Red dots represent data collected from

online databases (>50km proximity from reliable occurrence records) . Black line represents the Extent of Occurrence polygon from IUCN(<https://www.iucn.org/>).

1.3.2. Models and occurrence datasets evaluation

Evaluation metrics for UdRd models differed significantly from the rest (figure 1.3). This model had the lowest values for Sensibility, AUC, TSS and the highest values for Type II Error metrics. The unreliable dataset (UdRd) presented median values of 0.90 for Sensibility, 0.10 for Type II Error, 0.72 for TSS and 0.87 for AUC, whilst the FullSet and RdUd present values of 0.96 and 0.98 for Sensibility, 0.04 and 0.02 for Type II Error, 0.82 and 0.90 for TSS and 0.94 and 0.90 for AUC, respectively (figure 1.3). Moreover, the Kernel Density Estimation plot, represented as a violin plot in figure 1.3, shows that, for the FullSet and RdUd group, the higher percentage of the data are clustered around the median value whereas for the UdRd, the data are more scattered.

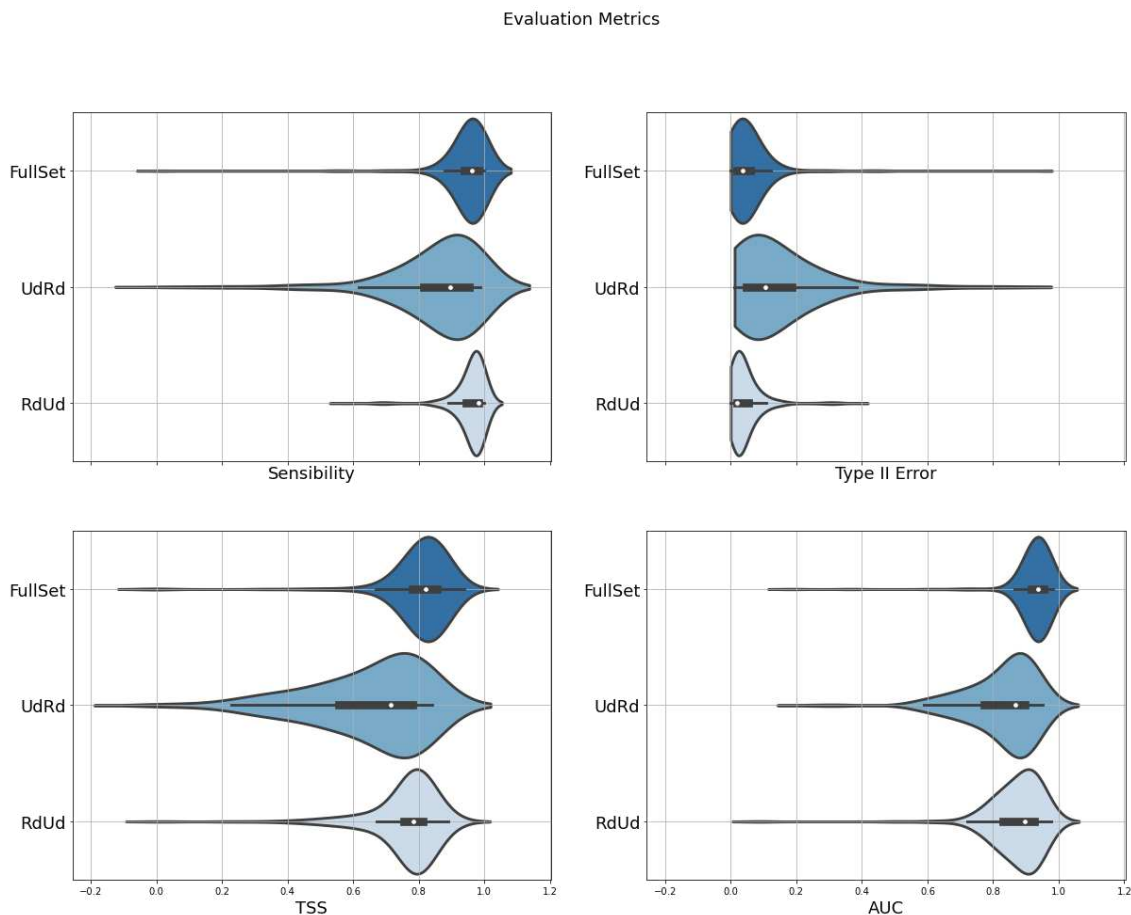


Figure 1.3: Evaluation metrics (Sensibility, Type II Error, TSS and AUC) for all three data groups (FullSet, UdRd,RdUd) presented as boxplots, indicating the median, as an white dot, and quartiles with whiskers reaching up to 1.5 times the interquartile range. Violin plot outlines indicate kernel probability density, i.e the width of the shaded area represents the proportion of the data located there.

The analysis of the uncertainty generated by modeling components (two-factor ANOVA - methods vs. data partitioning) revealed that data partitioning groups (PdRd, RdPd, FullSet) accounted for most variation in model performance statistics (Table 1.2, Figure 1.4). F values for two-factor ANOVA are higher for Data sets (FullSet, UdRd,RdUd) than for Methods (GLM, SVM, Maxent, BRT and RBF) (Table 1.2). Tukey test showed significant difference ($p < 0.001$) between UdRd and the other two groups (RdUd and FullSet) for Sensibility, Type II Error and TSS (Table 1.3). For AUC, results showed significant differences among all three groups (Table 1.3).

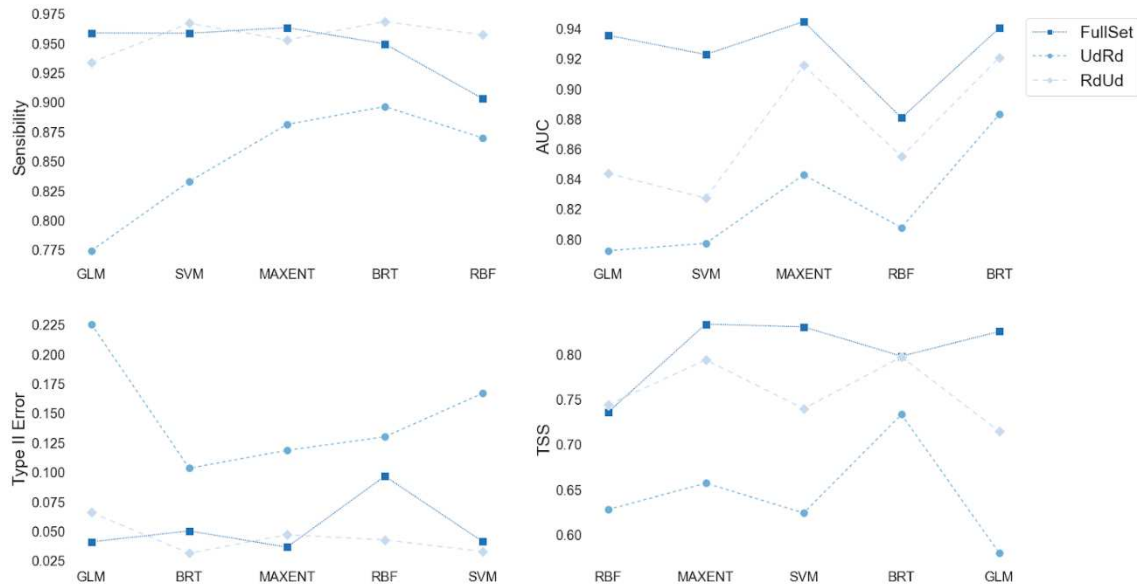


Figure 1.4: Line Plot of evaluation metrics (Sensibility, Type II Error, TSS and AUC) grouped by datasets (FullSet, UdRd,RdUd) for all algorithms (GLM, SVM, Maxent, BRT and RBF) used in the modeling exercise. Square markers represent the mean value for the FullSet dataset, diamond markers represent the mean value for the UdRd set and circle markers represent the mean value for the RdUd set.

Table 1.2: Two-factors factorial ANOVA results showing the effect of data partitioning (FullSet, RdPd, PdRd) and modeling methods (BRT, GLM, Maxent, RBF, SVM) on model performance metrics.

Source of Variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sensibility\$Algorithm	4	0.23	0.06	5.82	0.000129*
Sensibility\$Data group	2	1.69	0.84	86.10	< 0.000001*
Sensibility\$Algorithm*Sensibility\$Data group	8	0.41	0.05	5.24	0.000002*
Residuals	735	7.20	0.01		

TypeIIError\$Algorithm	4	0.23	0.06	5.82	0.000129*
TypeIIError\$Data group	2	1.69	0.84	86.10	< 0.000001*
TypeIIError\$Algorithm*TypeIIError\$Data group	8	0.41	0.05	5.25	0.000002*
Residuals	735	7.20	0.01		
TSS\$Algorithm	4	0.65	0.16	9.87	< 0.000001*
TSS\$Data group	2	3.42	17.09	104.36	< 0.000001*
TSS\$Algorithm:TSS\$Data group	8	0.61	0.076	4.66	0.000014*
Residuals	735	12.04	0.02		
AUC\$Algorithm	4	0.60	0.15	21.24	< 0.000001*
AUC\$Data group	2	1.26	0.63	89.41	< 0.000001*
AUC\$Algorithm:AUC\$Data group	8	0.20	0.025	3.52	0.000527*
Residuals	735	5.16	0.01		

* The mean is significant at the 0.001 level.

Table 1.3: Tukey Honest Significant Differences (HSD) *post-hoc* test results showing the effect of data partitioning (FullSet, RdPd, PdRd) on model performance metrics.

	Mean difference	P adjusted	95% confidence interval	
			lower	upper
TSS\$Data group				
UdRd-FullSet	-0.16	0.00000*	-0.19	-0.13

RdUd-FullSet	-0.03	0.001	-0.04	-0.020
RdUd-UdRd	0.13	0.00000*	0.09	0.14
Sensibility\$Data group				
UdRd-FullSet	-0.1	0.00*	-0.12	-0.07
RdUd-FullSet	0.01	0.55	-0.01	0.03
RdUd-UdRd	0.10	0.00*	0.084	0.13
TypeIIError\$Data group				
UdRd-FullSet	0.10	0.00*	0.07	0.12
RdUd-FullSet	-0.01	0.55	-0.03	0.01
RdUd-UdRd	-0.10	0.00*	-0.13	-0.08
AUC\$Data group				
UdRd-FullSet	-0.10	0.00*	-0.12	-0.08
RdUd-FullSet	-0.03	0.00*	-0.07	-0.03
RdUd-UdRd	0.07	0.00*	0.03	0.08

* The mean is significant at the 0.001 level.

Overall, in the environmental space, variable importance and response curve function analyses indicated that there are more similarities between RdUd and FullSet models compared to UdRd models (Figure 1.5 and Figure 1.6). Bathymetry was the environmental predictor that most contributed for all three models, followed by variables related to sea surface temperature for FullSet (biogeo14: sea temperature of the coldest month) and RdUd (biogeo15: sea temperature of the hottest month), and by mean annual salinity (biogeo08) for UdRd (Figure 1.5). Regarding the response curve functions, FullSet and RdUd models showed consistent curves with low confidence

intervals and similar results, with high suitability values in shallow waters, mean annual salinity between 30-36 PSU and low annual variation in salinity (Figure 1.6). Additionally, both models indicated high suitability values in warmer waters around 20°C in the coldest month and around 30°C in the hottest month (Figure 1.6). For the UdRd model, response curve functions presented inconsistent behavior, with curves showing high variability in the confidence intervals, meaning that the same value for a given environmental variable could indicate low or high suitability value in the consensus model (Figure 1.6).

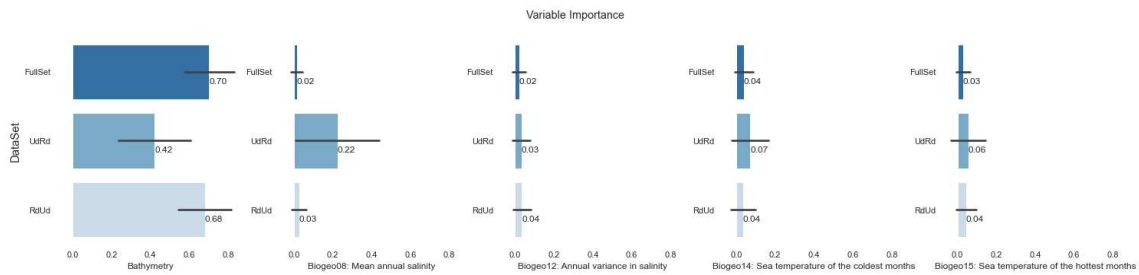


Figure 1.5: Contribution of the environmental variables to model performance measured by AUC. Values are the mean and standard of all 250 models generated for each occurrence dataset (FullSet, Rd, Pd).

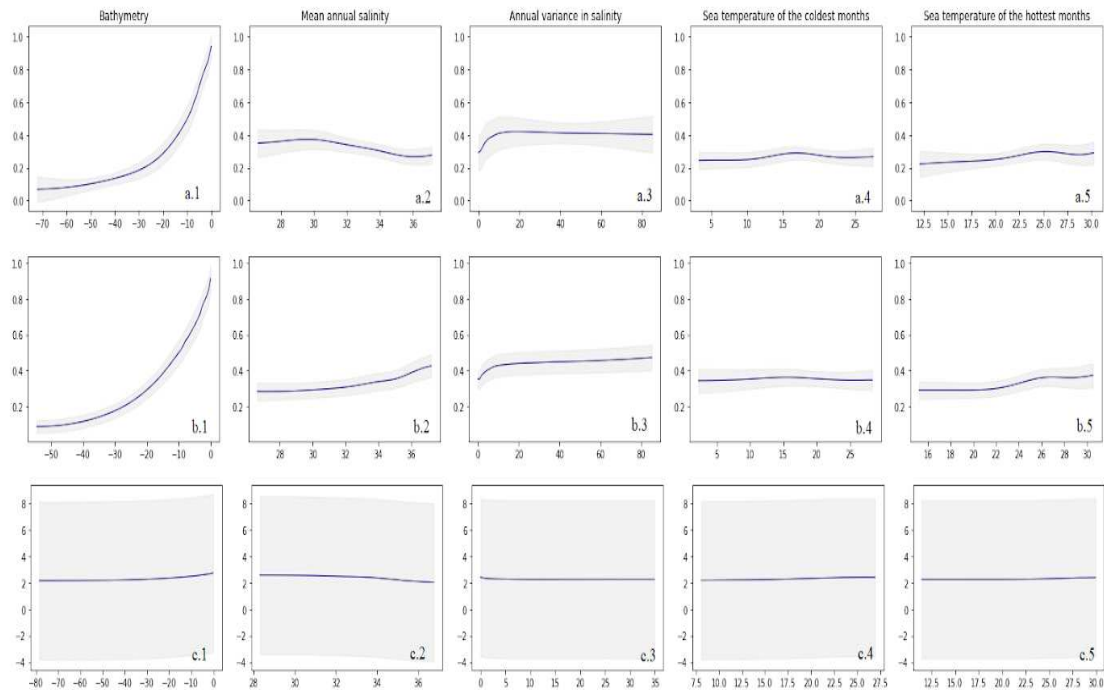


Figure 1.6: Response curves (black solid line) of the ensemble models generated for the three occurrence datasets (a1 to a5: FullSet model; b1 to b5: RdUd model; c1 to c5: UdRd). The gray shadow represents the confidence interval considering all 250 replicates for each of three occurrence datasets.

In the geographical space, comparison of Schoener's D revealed more similarity between RdUd and FullSet models than between each and UdRd models (Table 1.5).

Schoener's D calculated for the overlap of predicted suitable areas between the models calibrated with Rd and FullSet occurrence datasets was higher than 0.9, which indicates that overlap in geographically projected models is high.

Table 1.5: Similarity, in the geographic space, of the modeling outputs generated by the three-occurrence dataset (FullSet, RdPd, PdRd).

Models	Schoener's D	Spearman correlation
RdPd x FullSet	0.91	0.98
FullSet x PdRd	0.88	0.89
RdPd x PdRd	0.87	0.84

In terms of the distribution of pixel value for the models within the EOO of *Epinephelus itajara*, using the 0.5 suitability threshold, the FullSet and RdUd models produced more similar estimates. The distribution of pixel values, as a function of the probability density, are higher in high adequability bins for FullSet and RdUd models reaching its peak in the density of distribution between 0.8 - 0.9 of adequability (Figure 1.7 A and Figure 1.8 A). For the UdRd model, the higher density distribution of pixel values are clustered around the 0.5 - 0.6 adequability bins and start to decrease as the adequability increases. (Figure 1.9 A). Plotting these pixels in the geographic space, FullSet and RdUd had similar results. Both models predicted areas located all along the coast and with similarity to the known distribution of the species (Figure 1.7 B and Figure 1.8 B). The UdRd model presented large gaps between distribution and with only high values of suitability in the African portion of the distribution(Figure 1.9B). The total adequability values for the FullSet model was 3×10^6 km² , for RdUd 2.9×10^6 km² and for UdRd 2.3×10^6 km² .

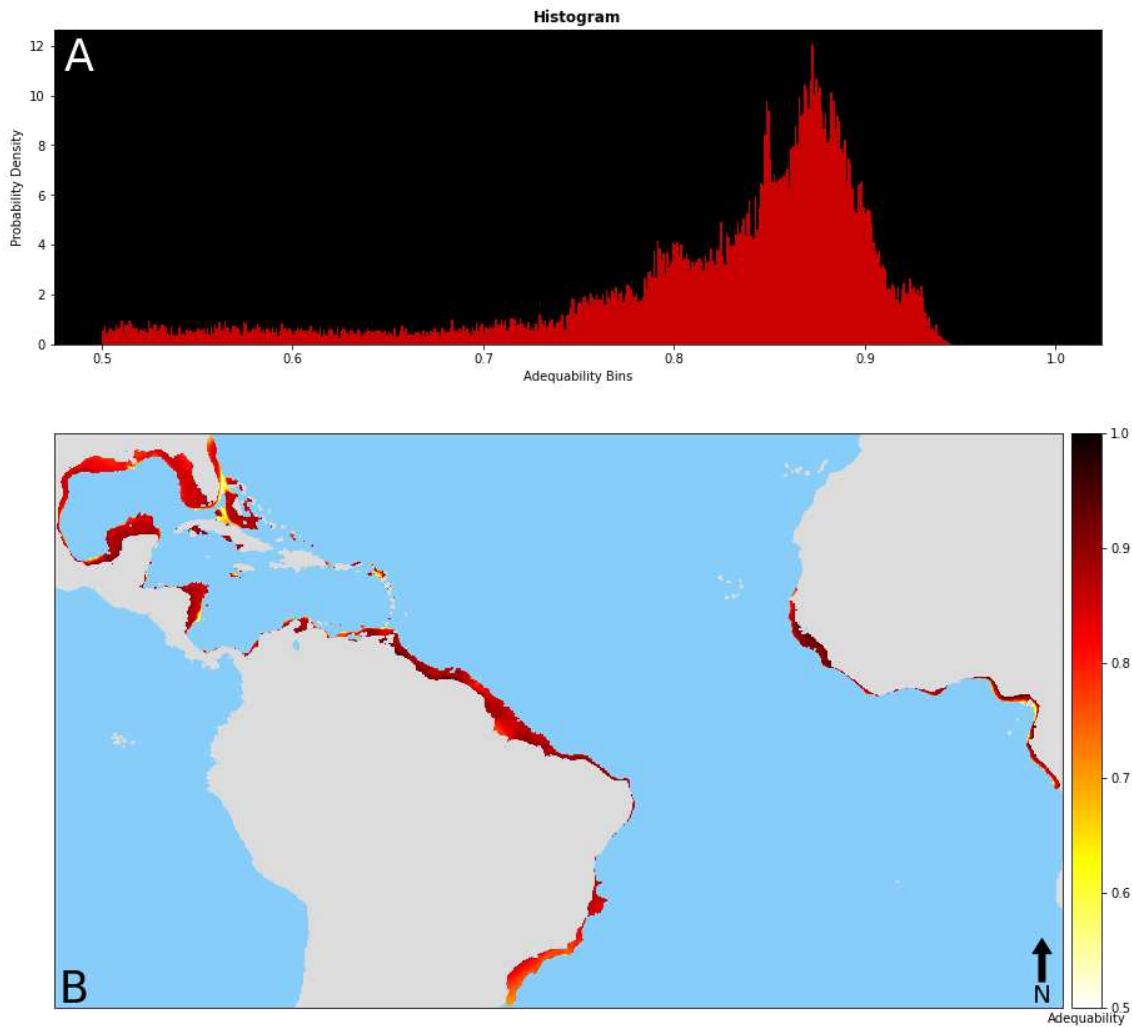


Figure 1.7: Distribution of probability density of pixel values across adequability bins for the FullSet model (A). B indicates the distribution of pixel values in the geographic space. Warmer red colors indicate high adequability values whereas yellowish colors indicate low adequability values.

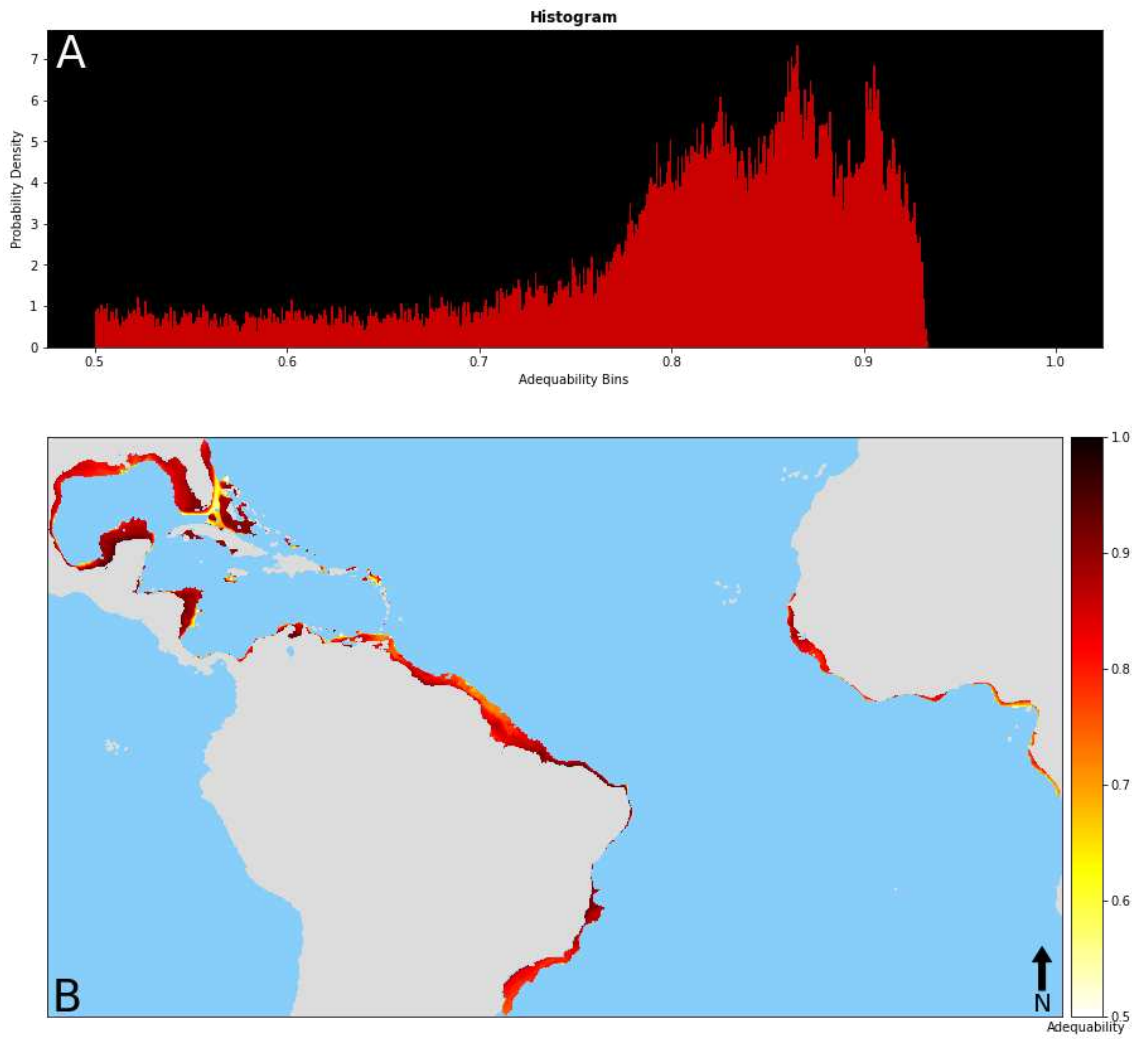


Figure 1.8: Distribution of probability density of pixel values across adequability bins for the RdUd model (A). B indicates the distribution of pixel values in the geographic space. Warmer red colors indicate high adequability values whereas yellowish colors indicate low adequability values.

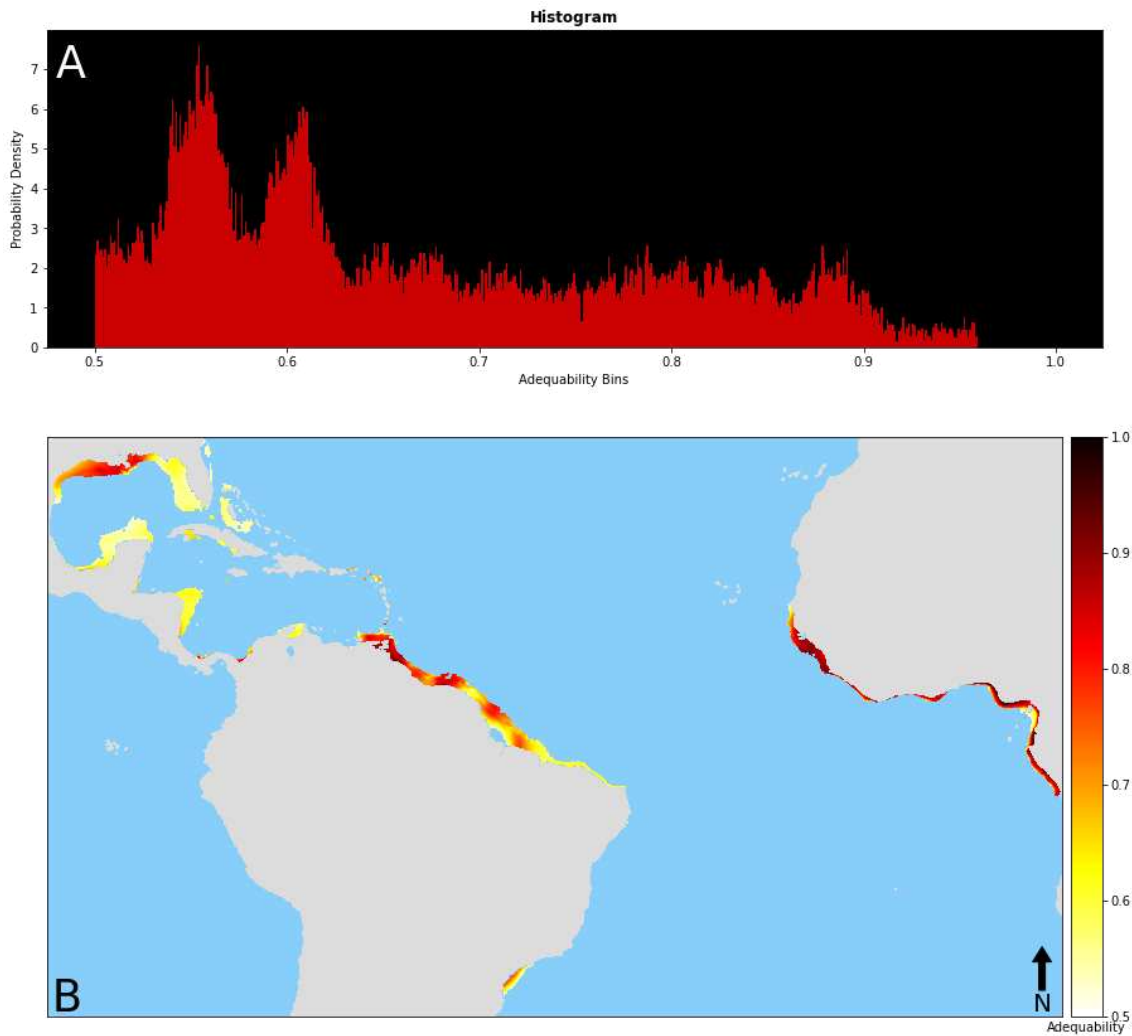


Fig 1.9: Distribution of probability density of pixel values across adequability bins for the UdRd model (A). B indicates the distribution of pixel values in the geographic space. Warmer red colors indicate high adequability values whereas yellowish colors indicate low adequability values.

The evaluation of the temporal coverage of the occurrence datasets revealed low coverage and a strong bias in Ud records. Time span ratio (occurrence datasets / all records) computed for Ud (0.04) was very lower than for Rd (0.91). While the Rd dataset presented a temporal distribution covering all the time span analyzed, the Ud dataset presented only old records restricted to the years of 1963 and 1964 (Figure 2.0).

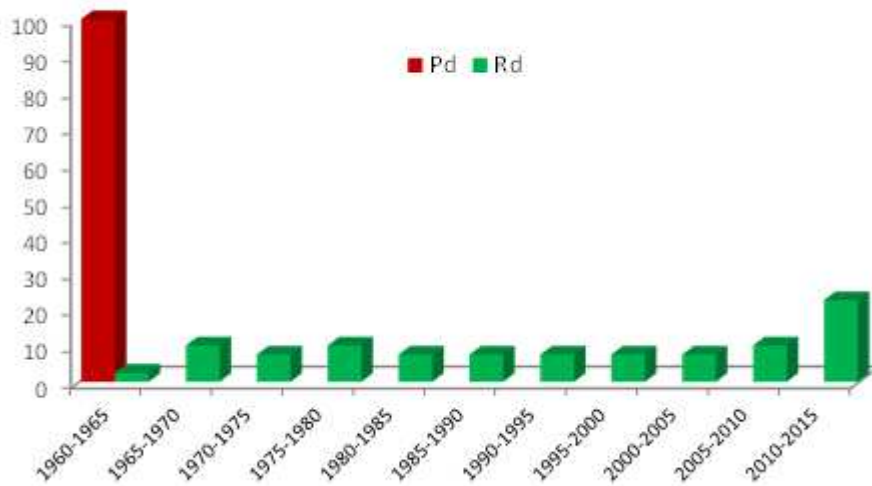


Figure 2.0: Temporal coverage of the occurrence datasets of *Epinephelus itajara* showing the percentage of unreliable (Ud) and reliable (Rd) records along the years.

The analysis of the environmental coverage of the occurrence datasets of *Epinephelus itajara* indicated that the environmental space occupied by the Ud records was not distinct that that occupied by Rd dataset. In the environmental space summarized by the first and second axes of the Principal Component Analysis (PC1 = 39.4% and PC2 = 32.4%), developed with the five variables used in the models, Ud environmental space appeared as a subset of Rd (Figure 2.1).

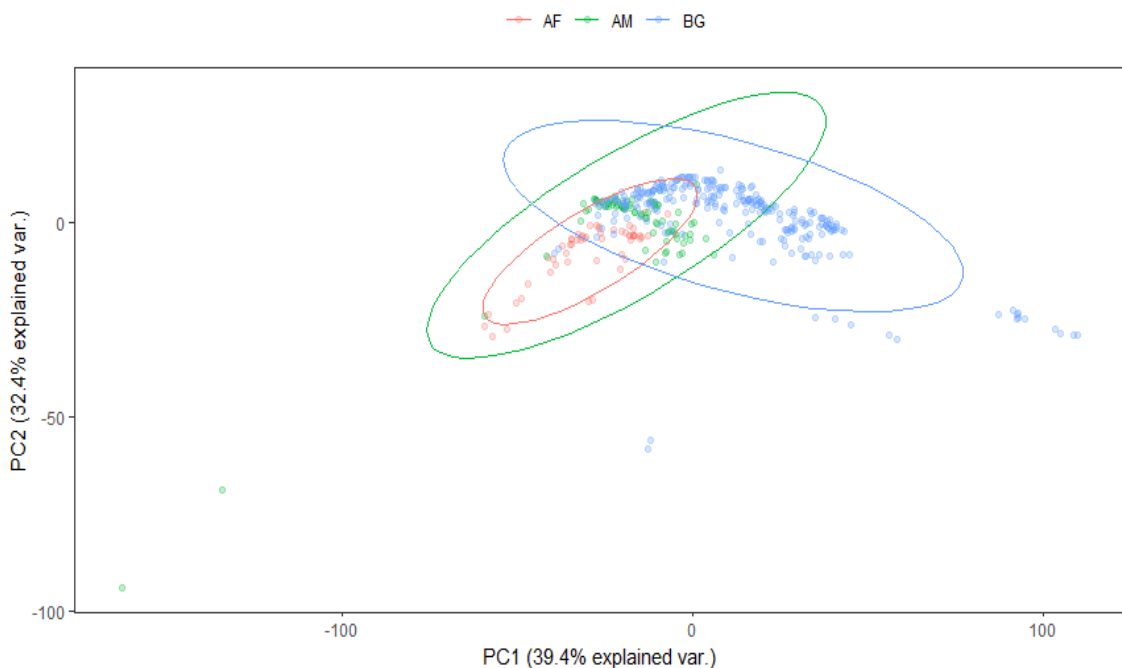


Figure 2.1: Environmental space of occurrence datasets of *Epinephelus itajara* summarized by the Principal Component Analysis composed by the five environmental variables used in the

models. Red dots - unreliable records (Ud). Green dots - reliable records (Rd). Blue dots - environmental conditions for the area accessible to the species, delimited by the ecoregions where the species is known to occur and adjacent, and sampled through 5000 randomly distributed points. Dots are enclosed by ellipses encompassing 95% of the data.

After the multiple aspects evaluated, we demonstrated that: (i) models fitted with data filtered to reduce uncertainty and biases (RdUd) outperformed those fitted with unfiltered data (UdRd), especially in Sensitivity and Omission Error; (ii) RdUd and FullSet models were more similar and more robust than UdRd in all performance metrics, as well as in variable contribution, response curves behavior and geographic prediction of suitable areas; (iii) UdRd models showed inconsistent response curves; (iv) Ud dataset represented a subset of Rd in the environmental space; and (v) Ud dataset presented low temporal coverage with only old records. Therefore, following the proposed workflow guidelines, the best decision to make is discard the Ud records and use the RdUd model results.

1.4.0 DISCUSSION

Our workflow, illustrated by *Epinephelus itajara*, yielded a framework to support decision about the use of unreliable records in ecological niche modeling, and shed light on how uncertainty in occurrence data can affect ENMs and overall measures of model accuracy.

Firstly, we demonstrated that more than one third (39.7%) of *Epinephelus itajara* occurrence records were assigned to very high uncertainty level. Similar finding was pointed in previous studies for mammal species in Australia (51.3%, Gueta & Carmel, 2016) and tree species at a global scale (48.5%, Jin & Yang, 2020). These results call the attention for the high proportion of low-quality records among data available from online occurrence repositories and the need for data cleaning in biodiversity studies (Maldonado et al., 2015; Gueta & Carmel, 2016; Zizka et al., 2019; Jin & Yang, 2020).

Subsequently, the rarefaction procedure applied to reduce spatial autocorrelation and sample bias filtered out a substantial portion (59.3%) of *Epinephelus itajara* occurrence records. Therefore, after these two filtering procedures, only approximately 1% of the occurrence records remained. It is impressive that from a total of 14,027 unique records, which could be considered a large dataset for a tropical species (Feeley

& Silman, 2011; Feeley et al., 2015), we retained only a small number of records ($n = 138$) to generate ENMs. An alternative could be applying environmental filters instead of geographic filtering to reduce spatial autocorrelation and sample bias. Although not widely used as geographic filtering, the use of environmental filters, derived from Principal Component Analysis axes, can be an advantageous rarefaction procedure. Castellanos et al. (2019) demonstrated that environmental filtering can return better performance in model evaluation statistics, show more biologically realistic predictions, and concomitantly retain more samples records than an equivalent geographic filter.

Following our workflow through the ecological niche modeling we were able to demonstrate that the PdRd model produced the worst result when compared to the rest of the models. According to the result of ANOVA and Tuckey SD, as well as of Schoener's D analysis, PdRd was the most dissimilar among the three models. This model also presented the biggest omission error. But regarding TSS, the PdRd model obtained a satisfactory evaluation. TSS is a statistic that ranges from -1 to 1 and positive values indicate models better than random results (Allouche et al., 2006). Models with $TSS < 0.4$ are considered bad predictions, $0.4 \leq TSS \leq 0.8$, useful and $TSS > 0.8$ from good to excellent (Zhang et al., 2015). In this sense, according to our results about TSS, the PdRd model should be classified as a useful model. However, all the other aspects of the PdRd model outputs, like the distribution of its predicted suitable areas, sensitivity value, omission error and the shape of response curves for environmental variables, indicated bad model performance. It is still possible to observe the low capacity of the PdRd model to predict suitable areas within regions of known occurrence the EOO of the species. Regions such as Florida, southern Brazil and the mouth of the Amazon River, which are known as nursery and seasonal aggregation for the Goliath grouper (Referencia IUCN e os artigos no docword) were considered as partially suitable or not suitable at all. Therefore, our findings indicated that TSS alone is not enough to discriminate between good and bad models. This is in accordance with recent studies that have criticized TSS, especially in the case of models based on presence-only or presence-background data, i.e. data with no information on locations where species do not occur (Jarnevich et al., 2017; Somodi et al., 2017; Leroy et al., 2018; Wunderlich et al., 2020). It was clear from our results that the use of occurrence records with some type of associated uncertainty can lead the model to have high omission error and low power to predict suitable areas in geographic space. This is line with the literature evidences demonstrating that the use of historic occurrence data at coarse resolution

decrease the performance of the models when compared to models trained only with recent records at a finer resolution (Reside et al., 2008; 2011). Only when the inclusion of data with uncertainty or geographically biased do not cause a bias in the environmental space the decision of using such data do not decrease model performance. This is the case pointed by Kadmon et al. (2004), when using biased records for highways, obtained models with reduced accuracy when compared to models corrected for this bias or with models that use records without this problem, although the magnitude of the reduction in accuracy was not statistically significant. This is not the case in our study. Clearly, through the PCA it is possible to observe that the records with some uncertainties are a subset, in the environmental space, of the reliable records set. Consequently, model calibrated with uncertain data (PdRd) probably would not be able to predict suitable areas outside that restricted environmental space.

We can conclude then those uncertain records decrease the model's performance, increasing their error of omission and decreasing their ability to project the models from the environmental space to the geographical space. We also emphasize that the exclusion of these uncertain and/or biased records would only be possible if the records with low uncertainty are enough to represent the entire environmental niche of the species. The decision about the quality of occurrence data to be used in ENMs is not trivial, but increasingly necessary. Evidently, there is a need to improve and expand the bases of occurrence records (Meyer 2016). But until then, practical guidelines like our proposed framework can be useful to support decision about the use of unreliable records, which can improve the quality of ecological niche models. This become especially relevant when the ENMs predictions are used to guide conservation and management use decisions.

Chapter 2: Adding habitat predictors to ecological niche models: an approach to improve model accuracy and suitability predictions

INTRODUCTION

The knowledge about species geographical distribution is fundamental to address a wide array of ecological, biogeographical and evolutionary questions, and underlies nearly every aspect of managing biodiversity (Franklin, 2010). Information on where species occur is crucial to support biodiversity conservation and sustainable use decisions and to guide appropriate actions. Unfortunately, the knowledge about species geographical distribution is often incomplete (Villero et al., 2017; Sofaer et al., 2019).

Ecological niche models (ENMs) provide a spatialized approach to produce consistent and repeatable information about species distribution, which can be very useful to inform decisions (Sofaer et al., 2019). ENMs are correlative models that use a variety of algorithms to explore the relationship between species occurrences and environmental variables. Once this relationship is determined, the model is used to estimate the species' niche in the environmental space. The niche model is then projected as a probability surface into a geographic space, representing the potential distribution of the suitable area for the species (Peterson et al., 2011; Bellard et al., 2012; Guisan et al., 2017).

The choice of environmental variables is fundamental for this kind of correlative models, and despite its proven and important effect on the predictions, has received little attention (Syphard & Franklin, 2009; Austin & Van Niel, 2011; Tulloch et al., 2016) and still remains a source of debate (Synes & Osborne, 2011; Bucklin et al 2015; Leitão & Santos, 2019). Due to the fact that climate is widely recognized as a major determinant of species' distributions (Brown & Gibson, 1983; Woodward, 1987), frequently environmental predictors for ENMs are selected from a set of climatic variables (Hageer et al., 2017). However, climate-only ENMs have been criticized and sometimes considered incomplete representations (Araujo & Peterson, 2012), because other factors may affect species distributions (Heikkinen et al., 2006; Chatfield et al., 2010; Austin & Van Niel, 2011). In response to those criticisms, there have been attempts to include additional non-climate predictors in correlative ENMs (Austin &

Van Niel, 2011), although there are comparatively few studies incorporating non-climatic environmental variables (Hageer et al., 2017). Temperature predictors are the most widely used and also the most important variables in both, terrestrial and marine studies (Bradie & Leung, 2017; Bosch et al., 2018; Mammola et al., 2020; Melo-Merino et al., 2020). Nevertheless, the non-climatic variables, when used, were often selected as important in modelling the species distribution (Bradie & Leung, 2017; Mammola et al., 2020), reinforcing the idea that including such predictors can improve ENMs predictions (Velazco et al., 2017).

Ecological niche models generally use abiotic variables such as temperature and precipitation as predictors for terrestrial environments, or ocean temperature and salinity as predictors for marine environments. Among non-climatic predictors, habitat variables are used very less often. Recent revisions (Bradie & Leung, 2017; Mammola et al., 2020) revealed that, in the terrestrial realm, temperature predictors were used in more than 75% of the studies, while habitat predictors were included in less than a third of them. Similarly, in marine studies sea temperature predictors were used in 73,2% of the models, but habitat variables were included in only a few studies (3,4%) (Melo-Merino et al., 2020).

In marine realm the presence or absence of coastal habitats, such as mangroves, marshes, seagrass and reefs, are essential for the distribution of species, considering that many fish and other organisms depend entirely or partially on these habitats (Feary, 2013). Such coastal environments are of extreme biological importance, supporting a great wealth of species (McHugh, 1976; Booth, 2018). Estuaries are environments of transition between continental waters and ocean waters; therefore, they have very specific physicochemical characteristics. In a simplistic way, the habitats that compose the estuaries are mangroves (in tropical regions), saltmarshes (in temperate regions) and even seagrass. Mangroves are home to a great diversity of flora and fauna (Macintosh & Ashton, 2002). Bacteria that reside in mangroves are an important agent for chemical control as in the control of the nitrogen cycle and in the decomposition of sulfates (Sherman et al., 1998). The aerial roots of mangroves are home to a diversity of diatoms and unicellular algae, which are important for maintaining high levels of the food chain (Robertson & Blaber, 1992). From the group of animals, mangroves support permanent or temporary residents. Many of the fish found in mangroves are juveniles, suggesting that this habitat serves as a nursery (Macintosh & Ashton 2002). In addition to being used as a nursery, fish use this habitat for food and protection (Sasekumar et al., 1992).

Mollusks and crustaceans are another group quite abundant in mangroves, having an important position in the trophic web and nutrient cycling (Robertson, 1986; Smith, 1987; Slim et al., 1997).

Seagrasses represent one of the richest and most important coastal habitats in the ocean and are a source of food for megaherbivores such as green turtles, dugongs and manatees (Orth, 2011). Among the ecosystem services, we can mention the high primary productivity, filtration of nutrients and contaminants, filtration of sediments and production and export of organic components (Green and Short, 2004). This environment also causes the damping of the waves, which allows the formation of nurseries for several species of fish, including species with high commercial value (Beck et al., 2001; Heck et al., 2003). Furthermore, the proximity of this environment to other habitats such as salt marshes, mangroves and coral reefs facilitates the trophic transfer and cross-use of these habitats by fish and vertebrates (Beck et al., 2001). This provides energy subsidies that are essential to maintaining an abundance of reef fish (Valentine & Heck, 2005).

Known as the rain forest of the oceans, coral reefs are home to a great diversity of marine species (Talbot, 1994; Beger, 2003; Hughes, 2002). Estimates of their diversity indicate between 600,000 and 9 million species that are reefs (Knowlton, 2008). In a simplified way, the dynamics of coral reefs, consists of coralline algae helping to cement large blocks of coral to each other making them form even larger blocks. Fish, sea urchins and other herbivores forage algae or seaweeds, making it possible to build new corals on the reef. In this way, coral reefs protect other animals from wave dynamics and / or predators and these animals help reefs to grow and maintain health (Maragos et al., 1996). In addition to this relationship that fish and other marine organisms have with coral reefs, these habitats are used to recruit new larvae that are often brought in from distant places, thus also serving as a nursery (Knowlton, 2008; Maragos; 1996).

Additionally, Robinson (2011) points out that for marine species, a static climatic perspective alone is not enough to build ecological niche models for these species. Ontogenetic changes, for example, are an important physiological trait that limits the distribution of marine species depending on their stage of life. In fact, this change in habitat use and environmental preference / tolerance is very common in marine species (Dahlgren & Eggleston, 2000; Wilson et al., 2008). The size of the geographic distribution for some species of pelagic fish varies according to the stage of

life, where larvae and juveniles have a smaller distribution and adults a wider distribution (Mullon et al., 2000). The degree of habitat specialization can increase or decrease with ontogeny (Beck, 1995; Halpern et al., 2005) and it is common for reef fish species to use habitats far from reefs (mangroves, seagrass and macroalgae) as nurseries (Dahlgren & Eggleston, 2000). Another feature, but this time behavioral that can limit the geographic distribution are the seasonal aggregations. The aggregations are for food acquisition, avoiding predation and mating (Ritz, 1994) and as in many ecological niche models it is assumed that the species - environment relationship is stationary and linear (Austin, 2002; Fortin et al., 2005) such behaviors that vary in time and space in the ocean are not taken into account.

Although sea temperature represents the most widely used predictor in ecological niche models for marine species (Melo-Merino, 2020), some studies have employed proxies to describe habitats used by species. Abecasis et al. (2014) used variables that describe the distance to certain substrates in addition to variables originating from bathymetry (slope, aspect and curvature) to build ecological niche models for benthic and demersal species. Pace et al. (2018) used depth, slope and Euclidean distance from the coastline when modeling the distribution of social groups of Sperm Whales with the premise that these variables are a proxy for environmental characteristics that affect the distribution of these whales and their prey. Alt et al. (2019) in a study with zoonotic marine parasites compared ecological niche models built only with environmental variables with models including environmental variables and a variable that is the distance to a nearest grid cell containing at least one definitive record of host instance. The models that took the host into consideration had their performance improved, either by the Area Under the ROC Curve (AUC), or by the evaluation of the spatial predictions generated by the models.

The use of proxies or distance variables for types of substrates or host seem promising but they do not take into account that the presence of these habitats and biotic interactions are not stationary. Estuarine habitats, for example, follow the level of the sea (Kennedy, 1990), therefore assuming that the relative distance from these habitats is stationary is dangerous, especially for vulnerable species that are dependent on habitats for various biological functions.

Here, we assessed the extent to which habitat variables, combining with climate and salinity, may increase the performance of ENMs for coastal species. We hypothesised that models calibrated with both climate/salinity and habitat variables will

have greater performance compared with models that incorporate only climate/salinity or habitat predictors. To test this hypothesis, we selected a particularly valuable model species, *Epinephelus itajara*, due to its dependence on different types of coastal habitats.

The Goliath grouper, *Epinephelus itajara*, is the largest grouper in the Atlantic, with historical distribution in Florida, in the south of the United States, along the Gulf of Mexico and the Caribbean Sea and along South America to Santa Catarina, in Brazil and in the east along West Africa, from Senegal to Angola. Being able to reach 2 meters in length and 400 kilograms, it is considered a fish associated with coral reefs and rocky shores, artificial reefs and oil platforms. Mangroves are the main habitat for juveniles, although there are records of juveniles living in seagrass, puddles and shallow rocky areas. Usually, *E. itajara* juveniles that reach approximately one meter in size, go to offshore regions like coral reefs. Its main form of reproduction is through seasonal aggregations where there are a large number of specimens in the aggregation sites, making it a species that is very popular for fishing. It recently went from Critically Endangered to Vulnerable by the IUCN Red List of Threatened Species due to species conservation efforts (Bertoncini et al., 2018). Despite this, fishing for this species is usually common and there is evidence of genetic population structure due to population decrease. There has been no record of Goliath grouper in Africa for at least nine years. To account for the species dependence on different types of habitats, we proposed a modeling approach in which variables describing the relative distance of suitable areas for some types of coastal habitats are used as predictors, and comparative analysis are performed between models with and without habitat variables. When considering the distance from the suitable area for these habitats instead from the habitat itself, we seek to mitigate the effect of stationarity.

In addition, as *E. itajara* is an iconic threatened fish species (Craig et al., 2009; Bertoncini et al., 2018) and coral reefs, rocky shores and estuaries are among the most impacted marine habitats (Halpern, 2008), we assessed: i) the exposure of Goliath grouper distribution to anthropogenic stressors (fishing, ocean pollution and population pressure), and ii) the level of protection coverage under the current Marine Protected Areas network.

2.0 MATERIALS AND METHODS

2.1 Occurrence data

Our occurrence records for *Epinephelus itajara* are from online database (GBIF, Fishbase, OBIS, SpeciesLink, Brazilian Biodiversity Information System), literature data (Artero et al., 2015; Creed et al., 2007 ; Damasceno et al., 2015; Tzadik et al., 2017; Zapelini et al., 2017) and data provided by fish and grouper specialists (see Chapter 1, section 2.1). In Chapter 1 we showed that *E. itajra* records with some kind of uncertainty increase the omission error, thereby decreasing the models predictive ability. Therefore, records classified with some uncertainty were excluded from our analysis.

Spatial autocorrelation may increase the uncertainties of niche models (F.Dormann et al., 2007) and some methods have been tested to mitigate these problems but with limited efficiency in presence-only models (El-Gabbas & Dormann, 2018a; 2018b; F. Dormann et al., 2007). As suggested in Chapter 1, to minimize sample biases and spatial autocorrelation, as well as retain more occurrence records, we have opted for spatial rarefaction that takes into account the climatic heterogeneity of the calibration area, through SDMTToolBox 2.4 (Brown, 2014). Specifically, the tool uses the first three principal components of the entire group of environmental variables to calculate climate heterogeneity. At the end of this process we obtained 92 unique and rarefied occurrence records suitable for modeling (Figure 2.1).



Figure 2.1: Spatial representation of the marine ecoregions (in light yellow) and the 92 presence records (green dots) selected for ecological niche modeling.

2.2 Models calibration area and predictors

Considering that the species occurs in shallow and coastal waters, the calibration area was based on the World Marine Ecoregions (SPALDING et al., 2007). We include only ecoregions in which species was present, plus those ecoregions immediately adjacent (see Appendix 1, Figure S1, Table S1), as some model evaluation statistics tend to increase when the model calibration area also increases (Barve et al., 2011) (Figure 2.1).

The selection of environmental variables was based on three main aspects: biological relevance to influence marine fish (Sandman et al., 2013; Bradie & Leung, 2016; Goodyear et al., 2017) and *E. itajara* distribution (Gilmore et al., 1978; Eklund & Schull, 2001; Frias-Torres, 2006; Koenig et al. 2007; Chapman et al. 2014; Shideler et al., 2015); reduction of collinearity of the variables of the original set; and greater contribution to the model. Twelve variables representing mean, amplitude, and limits of ocean surface temperature and ocean surface salinity, at a resolution of 5 arc-minutes (~10km), were downloaded from the BioOracle database (Assis et al., 2018). From the original set of environmental variables, we selected only those variables that contributed

most to the models according the Jackknife method and, less correlated, through the Spearman correlation (see details in Appendix 5, Table S.4, Figure S.5). At the end of the process we selected predictors that represent the average values of minimum salinity records (Salinity.Lt.Min), average records that represent minimum and maximum temperature records (Temperature.Lt.Min and Temperature.Lt.Max, respectively) (Table 2.1).

We created variables representing Euclidean distance from suitable areas for seagrass (Figure 2.2), mangrove (Figure 2.3), rocky reef (Figure 2.4), coral reef (Figure 2.5) and for the sum of all habitats (Figure 2.6) that were incorporated as habitat predictors in the models. Detailed description of the procedures used to generate the habitat predictors are provided in Appendix 6.

2.3 Ecological niche modeling

To generate ecological niche models, we use five algorithms that together cover a considerable range of different widely used modeling techniques: Generalized Linear Model (GLM), Boosted Regression Tree (BRT), Maximum Entropy (Maxent), Support Vector Machine (SVM), and Radial Basis Function (RBF). Considering that distinct algorithms can produce different results (Thuiller 2004, Lawler et al. 2006, Pearson et al. 2006, Diniz-Filho 2009, Buisson et al. 2010), to reduce uncertainties we used an ensemble approach (Araújo & New, 2007; Qiao et al., 2015), which combines results from different algorithms into a single consensus model.

Species-specific tuning of model settings has already proved to be efficient for improving the results of ecological niche models (Anderson and Gonzalez, 2011). Better parameters produce model outputs with low omission and commission errors (Pearson et al., 2006; Warren and Seifert, 2014). Regardless of the modelling algorithm, data partitioning and pseudo-absences strategies affect output and performance of models (Morgane Barbet-Massin et al., 2012; Radosavljevic & Anderson, 2014; Iturbide et al., 2015; Roberts et al., 2017). Guided to a sensitivity analysis conducted to define better parameterization for data partitioning and pseudo-absences (see details in Appendix 7), we used subsampling partitioning method with 90% of occurrence records as training and the 10% as test data, and a 1:1 pseudo-absence to presence ratio (92 pseudo-absences: 92 presence) (see Appendix 7, Table S6).

We created four sets of ecological niche models with different combinations of environmental (climate and salinity) and habitat (distance to mangrove, distance to seagrass, distance to rocky reef, distance to coral reef) variables: (i) a model with only environmental variables (hereafter EnvOnly), (ii) a model with environmental variables plus four separate habitat variables (hereafter EnvHabs), (iii) a model with environmental variables and the sum of all four habitats (hereafter EnvSumHabs), and (iv) a model with only habitats (hereafter HabOnly). All models followed the same parameterization described in the previous paragraphs. We also tested for correlation between environmental and habitat variables and between habitat variables through a Spearman correlation matrix. We exclude the most correlated variables that were least important for models. We evaluated variables importance through getVarImp function in the SDM package (Naimi and Araújo, 2016) that uses the Area Under the ROC Curve (AUC) value as the basis for indicating which variable is the most important. The variables retained for each model are listed in Table 2.1.

Table 2.1: Variables used to generate each ecological niche model. Salinity.Lt.min = Average minimum salinity records, Temperature Lt.max = Average maximum temperature records, Temperature.Lt.min = Average minimum temperature records, Dist.To.Mangrove = Euclidean distance variable for each presence cell predicted as suitable for the mangrove ENM, Dist.To.Rockyreef = Euclidean distance variable for each presence cell predicted as suitable for the Rockyreef ENM, Dist.To.Seagrass = Euclidean distance variable for each presence cell predicted as suitable for the Seagrass ENM, Dist.To.Any.Hab = variable that represents the distances to all habitats.

Model	Variable
EnvOnly	Salinity.Lt.min, Temperature Lt.max, Temperature.Lt.min
EnvHabs	Dist.To.Mangrove, Dist.To.Rockyreef, Dist.To.Seagrass, Salinity.Lt.min, Temperature.Lt.min
EnvSumHabs	Dist.To.Any.Hab, Salinity.Lt.min, Temperature.Lt.min
HabOnly	Dist.To.Mangrove, Dist.To.Rockyreef, Dist.To.Seagrass

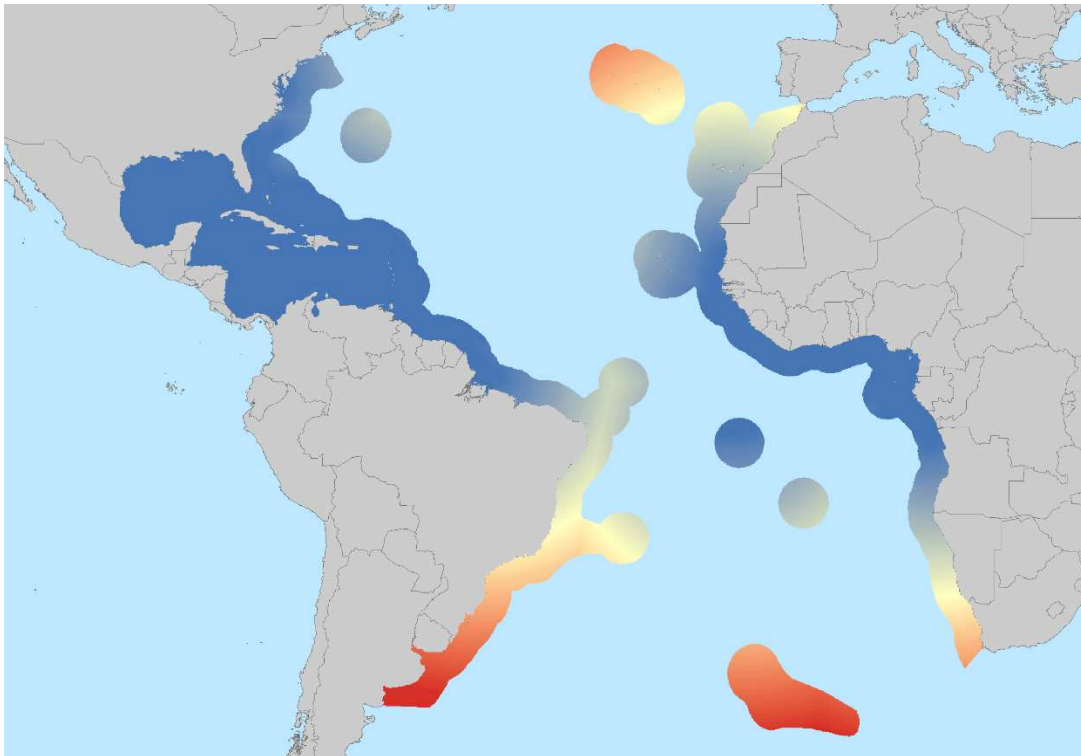


Figure 2.2: Euclidean distance variable for Seagrass habitat. Blueish colors indicate shorter distances to the appropriate seagrass areas and reddish colors indicate greater distances to the appropriate seagrass areas. The shortest distance found is 0 km and the largest distance is 4,973,768,000 km.

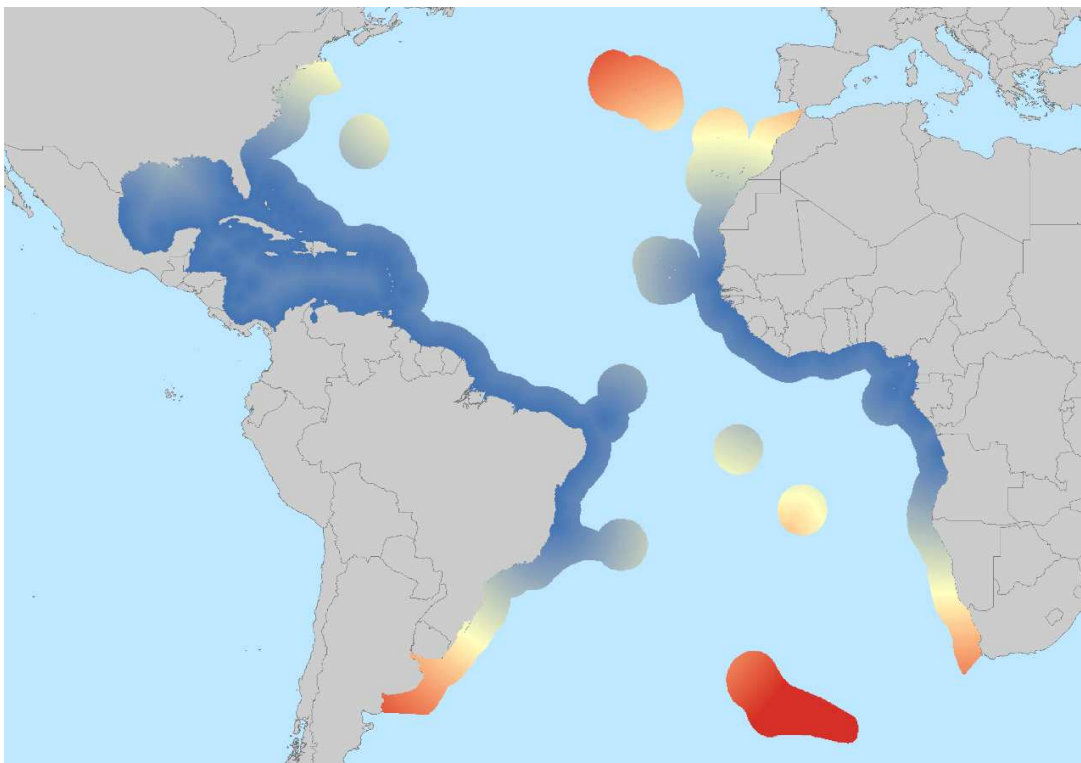


Figure 2.3: Euclidean distance variable for the Mangrove habitat. Blueish colors indicate shorter distances to the appropriate areas of seagrass and reddish colors indicate greater distances to the appropriate areas of mangrove. The shortest distance found is 0 km and the largest distance is 4,749,136,000 km.

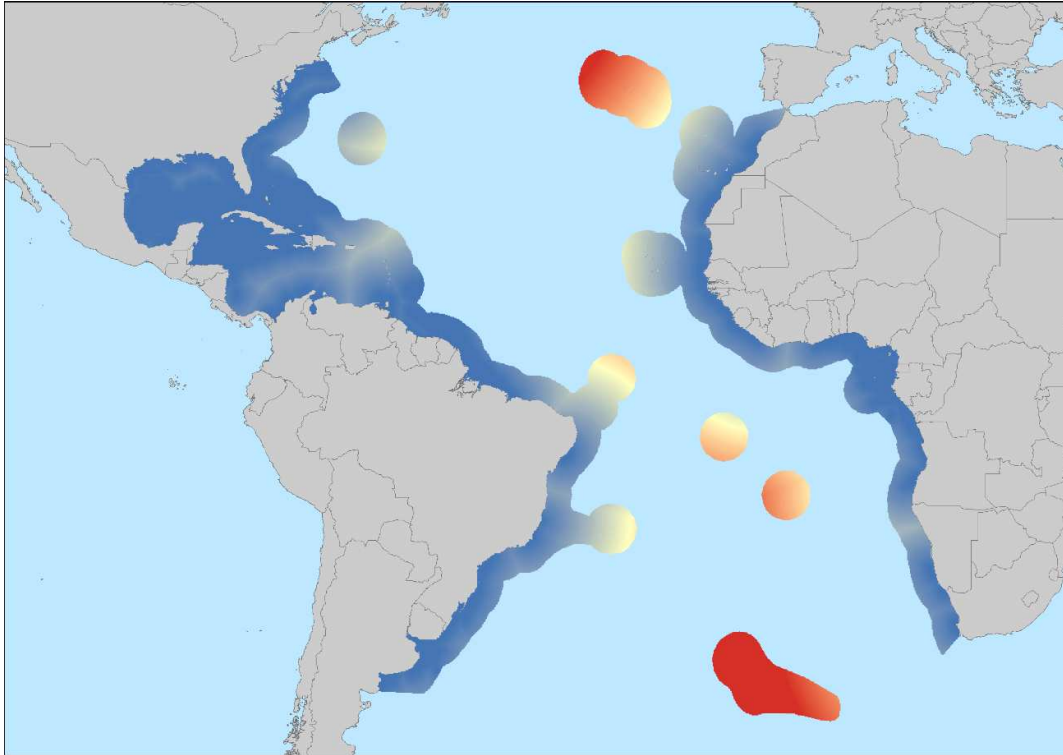


Figure 2.4: Euclidean distance variable for Rockyreef habitat. Blueish colors indicate shorter distances to the appropriate areas of rockyreef and reddish colors indicate greater distances to the appropriate areas of rockyreef. The shortest distance found is 0 km and the largest distance is 3,644,155,000 km.

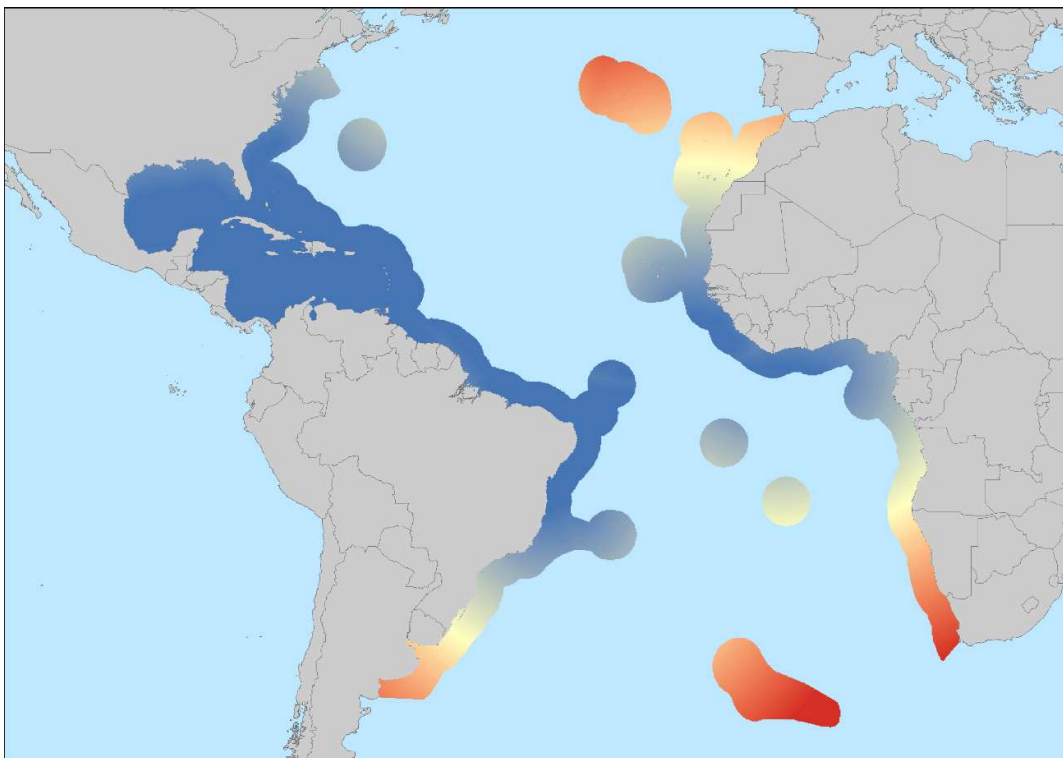


Figure 2.5: Euclidean distance variable for the Coralreef habitat. Blueish colors indicate shorter distances to the appropriate areas of coralreef and reddish colors indicate greater distances to the appropriate areas of coralreef. The shortest distance found is 0 km and the largest distance is 5,370,715,500 km.

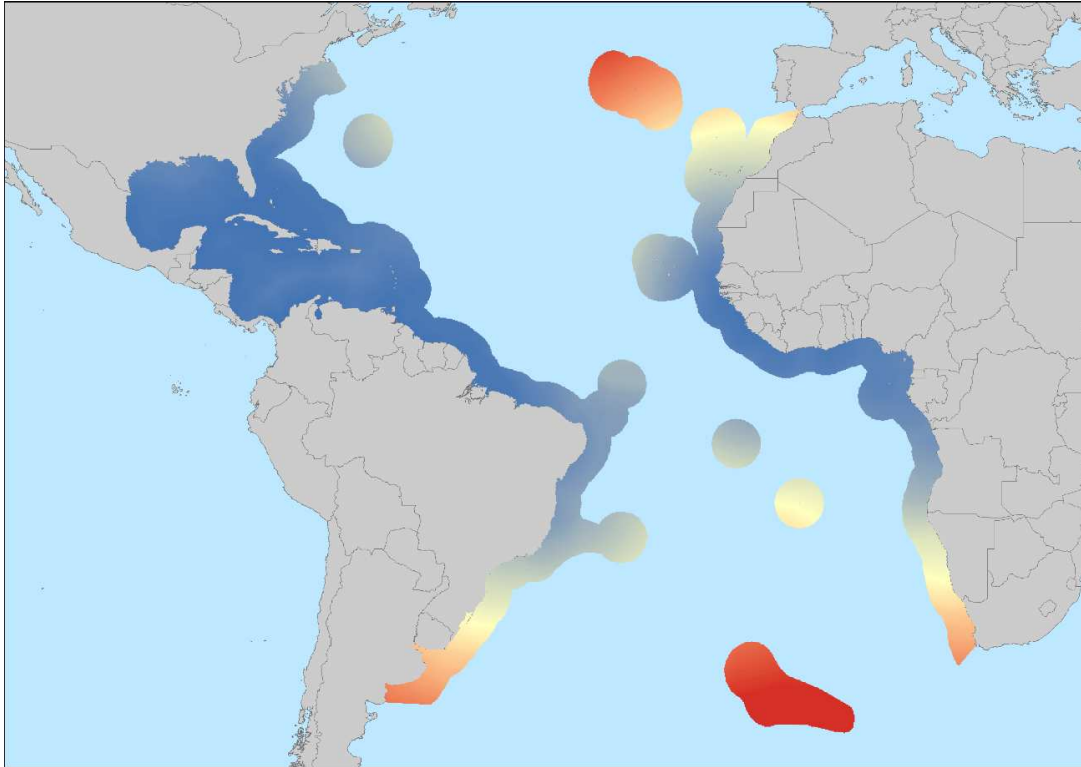


Figure 2.6: Euclidean distance variable for the sum of all habitats. Blueish colors indicate shorter distances to the appropriate areas of all habitats and reddish colors indicate greater distances to the appropriate areas of all habitats. The shortest distance found is 0 km and the longest distance is 16,913,706,000 km.

2.4. Models evaluation

The evaluation of the different ecological niche models was based on Sensitivity, Specificity and True Skill Statistics (TSS) generated from the confusion matrix. Sensitivity is the detected presence ratio predicted as such, so it quantifies omission error (Type II Error). Specificity is the detected absence ratio predicted as such, so it quantifies commission error (Type I Error). The TSS is calculated by the formula $\text{Sensitivity} + \text{Specificity} - 1$, that is, it takes into account errors of omission and commission (Allouche et al., 2006). Model performance and TSS values can be related as follows: poor (< 0.4), moderate ($0.4 > 0.6$), good ($0.6 - 0.75$), and very good (> 0.75) (Landis & Koch, 1977; Capinha et al., 2014). We built consensus ENMs using only good and very good models ($\text{TSS} \geq 0.70$), and applying TSS weighted average ensemble method (Araújo & New, 2007; Marmion et al., 2009). We used TSS because, although this metric has received some criticism as it is not immune to prevalence (Leroy, 2018) as previously thought (Allouche et al. 2006), it is still the most widely used metric to evaluate model's performance.

However, because we used presence-pseudo-absence models, we do not apply the TSS maximizing threshold (Liu, 2013), one of the most commonly used thresholds, to binarize our ensemble models. Instead, we use the Boyce Index (Boyce et al., 2002; Hirzel et al., 2006), a metric widely used lately to classify suitable areas (Cornellisen et al., 2018; Louppe, 2019; Hagar, 2020).

The continuous environmental suitability is reclassified into i number of classes. For each bin, Predicted and Expected frequencies are calculated. The Predicted Frequency is calculated by dividing the number of species' occurrence points in the bin i , as forecasted by the model, by the total number of species' occurrence points. The Expected Frequency is calculated by dividing the number of grid cells in bin i by the total number of grid cells. A P/E ratio is then calculated for each bin and a Spearman rank correlation coefficient ρ (1-tailed test) evaluates if the ratio significantly increases as suitability increases ($p < 0.05$). Models with good performance are visually identified by the P/E ratio curve graphic. The best P/E ratio curve is monotonically increasing (Boyce et al. 2002; Hirzel et al. 2006). Boyce's index interprets environmental suitability through quartiles, accounting for a threshold of unsuitability and then marginal to optimal suitability (Hirzel et al., 2006). The different quartiles and divisions of suitability's values will depend on the response curve of the model (Hirzel et al., 2006). The Boyce Index assesses how much of the model prediction matches the observed distribution of species occurrence through a predicted to expected ratio (P/E) curve. According to the shape of the P/E curve, P/E values < 1 were categorized as Not Suitable (NS) and P/E values > 1 were categorized as Suitable (S). The later included Moderately Suitable (MS) and Highly Suitable (HS) classes.

2.5. Distribution, extent, exposure to antropogenic stressors and protection of suitable areas

We analyzed the suitable areas predicted by the models in a pixel-by-pixel basis (Hu et al., 2010; Hu & Jiang, 2011) along ecoregions, under anthropogenic stressors and inside Marine Protected Areas (MPAs). Ecoregions were based on the World Marine Ecoregions (SPALDING et al., 2007) (see details in Appendix 1).

We considered only anthropogenic stressors that are pointed as relevant threats for *Epinephelus itajara* by the IUCN assessments (Craig et al., 2009; Bertocini et al., 2018). Fishing pressure is seen as the main factor for the decline of goliath grouper

population in the Atlantic (Bullock et al., 1992; Sadovy & Eklund, 1999; Zapelini et al., 2015) but other factors such as mangrove destruction (Sadovy & Eklund, 1999; Valiela et al., 2001; Koenig et al., 2007), mercury poisoning (Malinowski, 2019) and urban growth are also a threat (Craig et al., 2009; Bertocini et al., 2018). Accordingly, from Halpern (2015) we obtained data about four normalized variables that summarize these impacts. They are: (i) artisanal fishing, (ii) demersal, non-destructive, lowbycatch, (iii) ocean pollution, and (iv) population pressure (details about the generation of these variables can be seen in Halpern, 2015, supplementary material). In addition to these four variables, we used map algebra to generate a cumulative impact variable by summing all the impact rasters. Illustrations of these stressors' variables can be seen in the Appendix 8.

Using zonal statistics, we calculated the extensions and percentages of the Moderately Suitable (MS) and Highly Suitable (HS) areas for the species in the current network of MPAs. Data on MPAs (IUCN categories I-VI) were obtained from the Worldwide Database on Protected Areas (WDPA, <http://protectedplanet.net>).

Clipping procedures, spatial overlap, map algebra, zonal statistics and area calculations were performed in ArcMap. To avoid distortions in area calculations the data were converted to the World Equidistant Cylindrical Projection.

3.0 RESULTS

3.1. Model performance

The model generated with environmental variables (climate and salinity) in combination with distance variables for separate habitats (EnvHabs) reached the best performance (Sensitivity 0.904 ± 0.11 , Specificity 0.836 ± 0.13 , TSS 0.741 ± 0.15), followed by the model built only with distance variables for separate habitats (HabOnly) (Sensitivity 0.891 ± 0.11 , Specificity 0.818 ± 0.14 , TSS 0.710 ± 0.13). Models based only in environmental variables (EnvOnly) and models combining environmental with distance variables for all habitats (EnvSumHabs) had the worst performances (Table 2.2).

Table 2.2: Evaluation of the different ecological niche models for the Goliath grouper.

Model	Sensitivity	Specificity	TSS
EnvHabs	0.904 ± 0.11	0.836 ± 0.13	0.741 ± 0.15

HabOnly	0.891 ± 0.11	0.818 ± 0.14	0.710 ± 0.13
EnvOnly	0.769 ± 0.20	0.687 ± 0.21	0.456 ± 0.17
EnvSumHabs	0.830 ± 0.19	0.611 ± 0.22	0.44 ± 0.15

From the Boyce Index P/E Ratio curve for each model we can also discriminate between good and bad models (Figure 2.7). Good models should present increasing exponential curves and high P/E Ratio values (Hirzel et al, 2006). Boyce Index analysis pointed to EnvHabs and HabOnly models as the best models. In terms of curve shape, the EnvSumHabs, EnvHabs and HabOnly models showed increasing exponential curves, but only the last two presented high P/E Ratio (> 10). On the contrary, the EnvOnly model presented low P/E Ratio value and an exponential curve with a drop at its end.

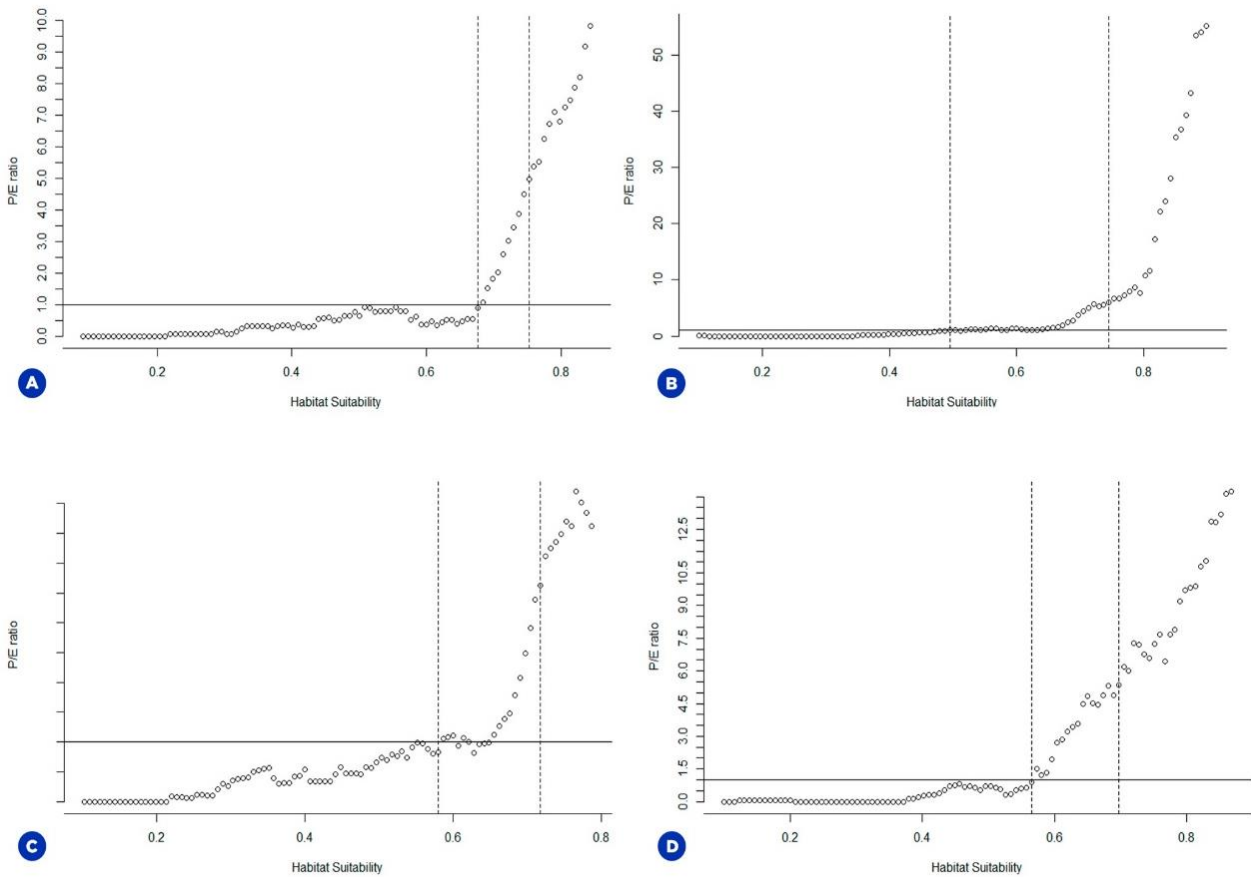


Figure 2.7: Boyce Index P/E Ratio curve for the four models: (A) Env+SumHabs, (B) Env+Habs, (C) EnvOnly, (D) HabOnly. For all plots, the horizontal straight line is where the

P/E ratio is equal to 1. The dashed lines are the limits for the classification between High Suitability (HS) and Moderate Suitable (MS), i.e., habitat suitability values before the first dashed line are Not Suitable (NS), values between the line are considered MS and values after the second line are considered HS.

3.2 Classification of suitable areas

According to the Boyce Index analysis, the P/E curve crossed the threshold of 1 at the suitability value of 0.48 for the EnvHabs model, 0.56 for HabOnly, 0.57 for EnvOnly, and 0.67 for EnvSumHabs. Pixels with suitability values ranging from 0 to these values were classified as Not Suitable (NS) areas and excluded from our analysis. Pixels classified as Moderately Suitable (MS) areas had suitability values between 0.48 - 0.74 for EnvHabs, 0.56 - 0.69 for HabOnly, 0.57 - 0.71 for EnvOnly, and 0.67 - 0.75 for EnvSumHabs model. Pixels with values higher than those mentioned above were classified as Highly Suitable (HS) areas (Figure 3.1).

3.3 Distribution, extent, exposure to anthropogenic stressors and protection of suitable areas

The model that predicted the larger extent of suitable area for *Epinephelus itajara* was the EnvOnly model, totaling $5,8 \times 10^6$ km² (MS = $3,7 \times 10^6$ km², HS = $2,1 \times 10^6$ km²) (Table 2.3). According to this model, suitable areas were not continuously distributed along the species' historical geographic range, showing a large suitability gap in the African coast, with the Brazilian coast partially suitable and with no suitable area in the Caribbean Sea (Figure 2.8C). The Env+Habs model predicted a total of $3,4 \times 10^6$ km² of suitable area (MS = $2,0 \times 10^6$ km², HS = $1,4 \times 10^6$ km²) (Table 2.3). The suitable areas according to this model were well distributed along the species' historical geographic range, with coastal regions generally showing a larger amount of HS areas and regions on the periphery of the species distribution or offshore showing a larger amount of MS areas (Figure 2.8B). The HabOnly model predicted a total of $3,1 \times 10^6$ km² of suitable areas (MS = $1,8 \times 10^6$ km², HS = $1,3 \times 10^6$ km²) (Table 2.3) and in geographic space, the model output showed good congruence with the known distribution for the species in Western Atlantic, covering the entire Brazilian coast, Caribbean Sea, and Gulf of Mexico (Figure 2.8D). The distribution of suitable areas in the African coast also showed good congruence with the known historical distribution of the species. The

Env+SumHabs model predicted the smaller extent of suitable areas, totalizing $2,2 \times 10^6$ km² (MS = $1,1 \times 10^6$ km², HS = $1,1 \times 10^6$ km²) (Table 2.3). The suitable areas were not continuously distributed along the species historical geographic range, with large gaps occurring on both sides of the Atlantic Ocean (Figure 2.8A).

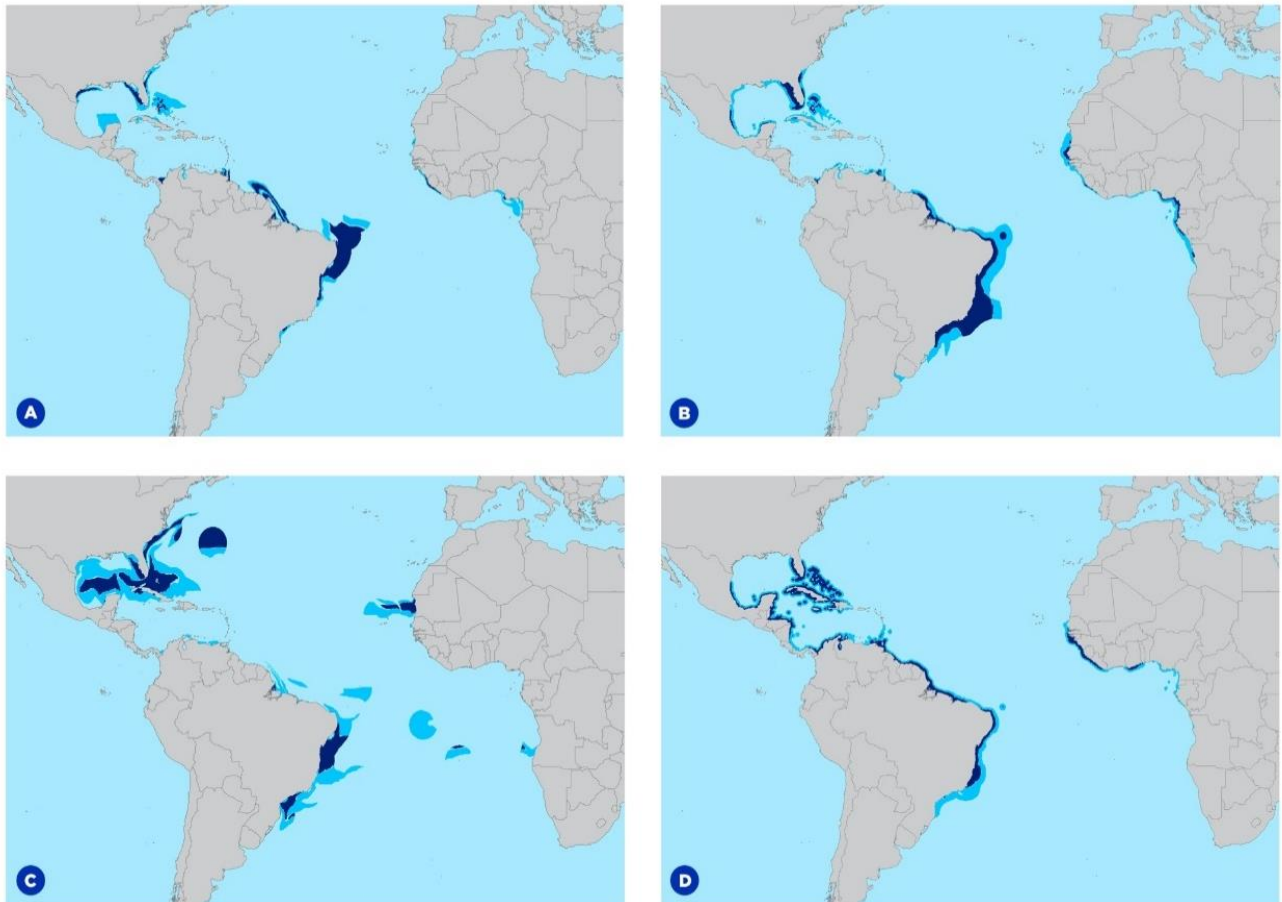


Figure 2.8: Environmental suitability for *Epinephelus itajara* predicted by the four ensemble models: (A) Env+SumHabs, (B) Env+Habs, (C) EnvOnly, (D) HabOnly. Dark blue colors represent High Suitability (HS) and light blue colors represent Moderate Suitability (MS).

In terms of human pressures, according to all models the suitable areas for *Epinephelus itajara* are under great exposure to all antropogenic stressors analyzed (artisanal fishing; demersal, non-destructive, lowbycatch fishing; ocean pollution and population pressure), showing similar values of impact. More than 95% of the pixels classified as suitable area in each model are under some type of human pressure (Table 2.4).

For all models, the results indicated that a low percentage (5,5 to 9,2%) of predicted suitable areas for *Epinephelus itajara* is under some kind of protection by the current MPAs network (Table 2.3). The model that predicted the higher amount (in km²)

of suitable areas within some protection level was the EnvOnly model, followed in decreasing order by HabOnly, EnvHabs and EnvSumHabs (Table 2.3). In percentage, the higher percentage was associated to HabOnly model, followed by EnvSumHabs, EnvHabs and EnvOnly (Table 2.3).

Table 2.3: Total of area predicted as suitable (S), highly suitable (HS) and moderately suitable (MS) for *Epinephelus itajara* and total area under Marine Protected Areas.

Model	S area (km ²)	HS area (km ²)	MS area (km ²)	S area inside MPAs (km ²)	S area inside MPAs (%)
EnvHabs	3,4x10 ⁶	1,4x10 ⁶	2,0x10 ⁶	2,1x10 ⁵	6,4
HabOnly	3,1x10 ⁶	1,3x10 ⁶	1,8x10 ⁶	2,8x10 ⁵	9,2
EnvOnly	5,8x10 ⁶	2,1x10 ⁶	3,7x10 ⁶	3,1x10 ⁵	5,5
EnvSumHabs	2,2x10 ⁶	1,1x10 ⁶	1,1x10 ⁶	1,4x10 ⁵	6,6

Table 2.4: Total predicted area within some pixel of human pressure.

Model	Artisanal Pressure (km ²)	Demersal Pressure (km ²)	Pollution Pressure (km ²)	Population Pressure (km ²)	Cumulative Pressure (km ²)
EnvHabs	3,36x10 ⁶	3,36x10 ⁶	3,36x10 ⁶	3,36x10 ⁶	3,34
HabOnly	3,0x10 ⁶	3,0x10 ⁶	3,0x10 ⁶	2,98x10 ⁶	2,97
EnvOnly	5,77x10 ⁶	5,77x10 ⁶	5,77x10 ⁶	5,77x10 ⁶	5,75
EnvSumHabs	2,2x10 ⁶	2,2x10 ⁶	2,2x10 ⁶	2,26x10 ⁶	2,25

4.0 DISCUSSION

Ecological niche models are frequently developed only with climate variables or climate and salinity variables for marine species, but does the addition of habitat predictors really improve model performance? Our results suggest that yes, habitat predictors in combination with climate have a strong influence on ENM accuracy and suitability predictions. Using *Epinephelus itajara* as a model species, we tested the hypothesis that models calibrated with both climate/salinity and habitat variables will have greater performance compared with models that incorporate only climate/salinity or habitat predictors. As we expected, the model that combined climate/salinity and habitat predictors for different habitats (EnvHabs) showed the best performance according all analysed metrics, Sensitivity, Specificity, TSS and Boyce Index.

ENMs are adjusted with variables to indirectly represent the physiological limits of a species (Peterson et al., 2011). Temperature and salinity are relevant variables to limit the distribution of marine species and their use has been extensive in ENM (Basher, 2016; Bosch et al., 2018; Baez, 2019; Alt et al., 2019; Catucci, 2020; Melo-Merino et al., 2020). Even in models calibrated with additional predictors, climate variables remain the more important predictors, suggesting that climate predictors have strong influence on ENM accuracy and predictions (Bucklin et al., 2015).

However, our results findings do not support this assumption proposed by Bucklin et al. (2015), and showed that models generated with only temperature/salinity variables (EnvOnly) are not enough to represent the niche of the target species. For the Goliath grouper, a species dependent on different types of coastal habitats (Craig, 2015), ENMs calibrated with climate/salinity and habitat variables (EnvHabs) and with only variables that represent the ecological optimal of the habitats (HabOnly) obtained the smallest omission and commission errors and highest TSS values (Table 2.2) when compared to the other models. From the Boyce Index P/E Ratio curve all models combining climate/salinity and habitat predictors showed increasing exponential curves, a characteristic of good models, while EnvOnly model, calibrated with only climate/salinity variables, presented low P/E Ratio values and an exponential curve with a drop at its end, which are characteristics of poor models. Decays in the P/E Ratio curve can be explained by a systematic bias in the observations or caused by biotic or environmental conditions that are not taken into account in the models (Cornelissen, 2019), however, we believe that this decay occurs because model does not differ from chance expectation or deviation from randomness (Hirzel, 2006) due to poor performance.

The choice of variables is an important decision in the construction of ecological niche models (Peterson et al., 2011). Criticism has been made about the reliability of such models, especially its biological relevance (Sofaer, 2019). To overcome this problem, the prior selection of variables of biological relevance is desirable (Foucard, 2019). There are studies that used variables as a proxy for habitats or biological interactions (e.g., vegetation cover, bottom type of marine substrate), obtaining satisfactory results (D. Space 2018 and review in Sofaer 2019) with models even better than those calibrated with only climatic variables (Alt et al., 2019). Ecological niche models can be used as a tool for biodiversity conservation (Guisan et al., 2013), but their results must be accurate so that the action to be taken is the most correct. The

literature shows us that conservation works that account for biotic interaction (Fordham et al., 2013) or that use proxies for habitat (Abecasis, 2014) have obtained satisfactory results and are faithful to the distribution of the target species.

Our results also pointed out that ecological niche models that include habitat variables (EnvHabs and HabOnly) are able to predict in more detail the biogeographic patterns and the historical distribution of the species. For both models the predicted HS areas were identified next to the coast, while MS areas were located far from the coast, which corroborates the literature that characterize Goliath grouper as a coastal species (Craig et al., 2009; Bertoncini et al., 2018). Additionally, for both models, there were large extents of suitable areas in part of the coasts of Brazil and Florida in the United States, mainly for the EnvHabs model. This can be explained because there is a high concentration of occurrence records in these two regions, which are famous for being nurseries and aggregation regions for mating, including Parque dos Meros in Brazil (Giglio, 2014) and Ten Thousand Islands (Lara, 2009; Tzadik, 2017). Nevertheless, both models had some peculiarities: the EnvHabs model presented some gaps in the distribution of suitable areas, especially for the coast between Venezuela and Brazil, Panama and Belize and a small extension of the American coast bathed by the Gulf of Mexico. For the HabOnly model, there was a gap in a larger portion of the American coast bordering the Gulf of Mexico and a breakdown in the distribution of suitable areas in southern Brazil. This gap between Brazil and Venezuela can be explained by the Orinoco - Amazonas plume which is a famous biogeographic barrier for marine animals (Briggs, 1974; Floeter, 2008). The gap between Panama and Belize can be explained by the low sampling of occurrence records, considering the results of the Boyce Index analysis. The region where there was a gap in the American portion of the Gulf of Mexico overlaps with the Mississippi Delta. This Delta has an important role as a biogeographic barrier (Brant & Orti, 2003; Pyron & Burbrink, 2009), which can also explain the gap presented by the HabOnly model, but in larger proportions. Likewise, the same pattern of gaps on the African coast can be explained by the occurrence of the mouth of several rivers. The gap in southern Brazil predicted by the HabOnly model occurred just below the Capricorn Tropic. This finding is in line with the limits of the distribution of suitable areas for our mangrove and seagrass habitat ENMs. Seagrass occurrence is reported to be low on the Brazilian coast due to low sampling or to absence of species due to unfavorable oceanographic aspects, while the highest concentration of seagrass beds is located in the Gulf of Mexico and the Caribbean Sea.

For mangroves, it is known that the distribution of this ecosystem is confined to tropical and subtropical regions, with low temperature in high latitudes being a limiting factor for the distribution of the species (Giri, 2010). The distance to rockyreef variable did not suffer this latitudinal effect as the entire shoreline of the calibration area was considered as habitat. It is noteworthy that both models correctly predicted the Amazon Delta as a suitable area, in line with the literature highlighting that the region functions as a nursery for the species (Lobato, 2015). One last difference we can report in relation to the distribution of the suitable areas predicted by the two models (EnvHabs and HabOnly) is that adding variables representing oceanographic aspects of temperature and salinity increased the model capacity to predict large areas of high suitability (HS) beyond the coast in certain regions, particularly Brazil and Florida. We do not consider that this expansion can be a Type I (Commission Error) error, since the models had high specificity (>0.8) (Table 2.2). Rather, this expansion can be a reflect of adding more information to the model, turning it more able to predict in greater detail the suitable areas for the species.

Our results about exposure to anthropogenic stressors for suitable areas predicted by both best models (EnvHabs and HabOnly) showed us a worrying situation for the target species. Goliath grouper is a species that had the conservation status modified from Critically Endangered to Vulnerable, according to the latest IUCN assessment. This is mainly due to conservation actions and the prohibition of fishing (Sadovy & Eklund, 1999; Gerhardinger; 2009). Fishing has always been an activity with high impact on marine megafauna (Jackson et al., 2001) and with Goliath grouper is no different. The use of longline and pearfishing is reported as one of the main forms of hunting the species (Bullock et al., 1992; Sadovy & Eklund, 1999; Zapelini et al., 2015) and our results support these findings. For EnvHabs model, 99% of all suitable areas are exposed to artisanal or demersal fishing pressure. For the HabOnly model, 96% and 97% of all suitable areas are under artisanal and demersal fishing pressure, respectively. Evidence shows that even with the prohibition of fishing, the activity is still carried out, reaching as many as 12 tons of the species in a decade in Brazil (Giglio, 2014). In addition, Goliath grouper is a curious species in front of divers (Sadovy & Eklund, 1999) turning it even more susceptible to spearhunting fishers. Our results also showed that suitable areas are highly exposed to anthropogenic stressors of ocean pollution and human pressures (urban engineering, port activity, waste disposal) with at least 96% of these areas under impact. These two pressures can jeopardize the species

through impacts on its feed and habitat resources. *Epinephelus itajara* commonly feeds on invertebrates and benthic fish with low movement capacity (Koenig & Coleman, 2009; Tzadik et al., 2015; Malinowski et al., 2019) and such items on the species menu are a source of great concentration of mercury (Malinowski et al., 2019) making our target species susceptible to contamination through trophic biomagnification. Coastal habitats are globally susceptible to terrestrial and marine pressures (Halpern, 2008). Mangroves are an important habitat for the species, whose main function is serving as nursery, protection and feeding place for juveniles (Frias-Torres, 2006; Koenig et al., 2007), but they have been suffering from deforestation for a long time (Spalding et al., 1997; Spiers, 1999, Murray et al., 2003, Giri et al., 2011) and what is left over is often in degraded conditions (UNEP, 2004; MAP, 2005). Other impacts such as climate change (Ward, 2016) and even hurricanes (Graham, 2009) have an expressive role and can change the number of individuals in the Goliath grouper populations. The same scenario of a degraded environment is repeated for seagrass. Population growth and water contamination are among the main factors for the deforestation of seagrass meadows (Kemp et al., 1983; Larkum & West, 1990; Peters et al., 1997), and seagrass from estuarine areas is more vulnerable to contaminants from anthropogenic sources than those from the coast (Green and Short, 2004). Although not used in our models, coral reefs are also an important habitat for the target species (Frias-Torres, 2006; Koenig et al., 2007), which unfortunately is also suffering from high anthropogenic impacts.

Likewise, our results about protection level for suitable areas predicted by both best models (EnvHabs and HabOnly) raise concern about the conservation of the target species. Both models indicated that only a small extent of suitable areas (<10% of total suitable area) were within some MPA. The current MPA network seems to more effective at protecting HS than MS areas (Table 2.3), as most MPAs cover the coast more than non-coastal environments in the study area.

To conclude, ecological niche models represent a very useful tool to understand species distribution patterns, as well as and the exposure of their areas suitable to anthropogenic stressors and the coverage under protection areas, and it is important to implement such a tool in conservation action plans. If the focus of the work is with one or a few species, we suggest caution in the construction of the models, as we prove that, depending on the case (especially if the species has a strong association with certain habitats or environments), climatic variables alone are not enough to represent all the

niche of the species. Ecological niche modeling is not in its infancy, but there is an urgent need to report in detail its methodology regarding occurrence records, explanatory variables, model parameterization and evaluation statistics (Araújo et al., 2019; Zurell et al., 2020). We also highlight that more care must be taken when choosing the predictors to calibrate the models, and suggest that including additional non-climate variables, particularly habitat predictors, should be recommended. Although this can be challenging in many cases due to data unavailability, time and/or resource limitations, or incomplete ecological knowledge about predictors (Bucklin et al., 2015; Fournier et al., 2017; Velazco et al., 2017), it seems to be very promising.

Chapter 3: Ecological niche models predict, in a scenario of future climate change, a shift to high latitudes for a vulnerable species of grouper.

3.1 INTRODUCTION

It is increasingly clear that climate changes driven by human activities are part of our lives (Keller, 2007; IPCC, 2104; 2018) and this problem has received more acceptance and attention from the mainstream media and governments (Oreskes, 2004; Boykoff & Pearman, 2019). If greenhouse gas emissions continue to rise at the current rates, predictions indicate that global mean temperature can increase beyond 4°C above pre-industrial levels until the end of XXI century, with high impacts on human and natural systems (IPCC, 2104; 2018; Días et al., 2019). It is undeniable that anthropogenic activities have been causing changes in the Earth climate and, consequently, in its biota (Días et al., 2019). Temperature increases, especially beyond 1.5°C or more, a threshold predicted to be reached between 2030 and 2052, increase the risk associated with profound or irreversible changes, such as the loss of some ecosystems (IPCC, 2018).

The increase in temperature directly affects the precipitation, drainage and constituents of the hydrological cycle (Klige, 1990; Zestser and Loaiciga, 1993; Loaiciga et al., 1996). Floods, precarious conditions for capturing water and excess underground runoff are some of the problems that arise from an imbalance in these hydrological cycles on the planet Earth (Loaiciga et al., 1996; Milly et al., 2002). About 6% of the influx of water into the oceans comes from groundwater (Zestser and Loaiciga, 1993) so an increase from these waters directly affects the volume of water in the seas. In addition to this increase caused by the greatest influx from the continent, the increase in temperature in the oceans causes the volume to also increase due to a thermal expansion of the waters, contributing even more to this increase in volume (Stevenson et al., 2002).

The oceans are responsible for absorbing more than 90% of the excess heat that remains trapped in the Earth's climate system from the 1970s to the present day, making them a key component of the planet's energy balance (Bindoff et al., 2013). There are observational evidence for changes in global ocean heat content (e.g. Barnett et al.

2001, Reichert et al. 2002) and ocean temperature and salinity (e.g. Barnett et al., 2005; Palmer et al., 2009; Pierce et al., 2012; Gleckler et al., 2012; Bilbao et al., 2019; Tokarska et al., 2019; Silvy et al., 2020). Such changes derived from human greenhouse gas emissions have effects on several marine systems aspects like stratification, oxygen concentrations, pH, primary productivity, circulation (Hartmann et al., 2013; Rhein et al., 2013; Pinsky et al., 2020). Recent observations indicate that oceans have experienced rapid warming over the last decades, contributing to increases in rainfall, sea levels; to decreases in oxygen levels, ice and glaciers; and to destruction of coral reefs (Rhein et al. 2013; Chen et al., 2019). Estimates indicate that 20 to 50% of the Atlantic, Pacific and Indian basins may already be affected by climate change, with models predicting that this proportion may reach 40–65% in 2050 and 55–80% in 2080 (Silvy et al., 2020). These changes can have catastrophic consequences for marine biota (Roessig, 2004; Boot, 2018).

Species can cope with climate change by shifting along one or several of three distinct, but not excludent axes (Bellard et al., 2012). In the spatial axis, species can move to regions where there are appropriate conditions to survive. In the temporal axis, species can change their rhythms and phenology. In the self axis, species can change life history traits in its physiology to cope with new climatic conditions. The latter response is more difficult to observe and with little documented evidence (Parmesan, 2000; Bellard et al., 2012). Evidence from the beginning of the century (~ 1910 - 1940) in marine environments showed that species contract or expand their geographic distribution due to the warming of the seas. Boreal fishes expanded their distributions to higher latitudes in the North Atlantic, while some cold-water fishes had their distribution contracted polewards (Drinkwater, 2006; Sundby & Nakken, 2008 Phillips and Pérez-Ramírez, 2018). This same pattern is repeated for climate changes in the modern period (1970 onwards). In a meta-analysis on the biological response of marine species to climate change, Poloczanska et al. (2013) reported that changes in geographical extension are already happening, especially in high-latitude regions, resulting in a reconfiguration of ecological communities, in ways consistent with theoretical expectations. Publications on changes in species distribution driven by anthropogenic warming are mostly focused on the eastern portion of the North Atlantic Ocean, and on temperate species (Poloczanska et al., 2013), but evidence with tropical reef fish is starting to appear (Feary, 2013). Climate change is already driving poleward range edges of marine species to expand at an average of 72 km/decade, which is

approximately an order of magnitude faster than observed rates on land (Poloczanska et al., 2013). Existing records have been sufficient to document hundreds of species moving to higher latitudes (Poloczanska et al., 2013; 2016; Pinsky et al., 2020). A recent study using a physiological non-spatial approach and including future climate changes showed that marine species can be more vulnerable to temperature rise than terrestrial species (Pinsky et al., 2019). Marine species may be more affected by climate change as the ambient temperature controls their geographic distribution, the availability of nutrients and the availability of oxygen in the ocean. Climate change is already altering the occurrence of marine species worldwide, reorganizing what has historically been considered the native and usual distribution of species (Pinsky et al., 2019; 2020).

Regions in Florida (USA), which experienced an increase in temperature, also experienced changes in their fish assemblage, with an increase in tropical or subtropical species that were not previously found (Fodrie et al., 2010). In southeastern Australia, Last et al. (2011) pointed out that 45 species of warm water fish colonized or at least increased their geographical distribution to colder regions, due to an increase in sea temperature and the strengthening of the Eastern Australian Current. In addition, at a 2011 La Niña event in Western Australia, Wernberg et al. (2012) recorded a 20% increase in the number of tropical species in the fish community. The increase in temperature in South Africa also contributed to decrease the number of seasoned fish caught, with a consequent increase in the number of tropical fish caught (groupers and sea bream) (Lloyd et al., 2012). Scientific evidence from the past demonstrates that changes in the distribution of marine fish species following environmental changes often occur, and can help to make some predictions of how future anthropogenic changes may affect the distribution of species.

As climate change represent one of the main global threat to biodiversity, being the most ubiquitous anthropogenic stressor (Días et al., 2019), assessing the species exposure and vulnerability to climate change become crucial. In the last decades, Ecological Niche Modeling (ENM) has become one of the most used approach to assess species exposure to climate change (Peterson et al., 2011; Russo et al., 2016; Araújo *et al.* 2019). Despite its limitations, mainly because it does not consider biotic interactions (Araújo & Luoto, 2007; Gaston & Fuller, 2009), the contribution of this tool to the understanding of climate change impacts on biodiversity is undeniable (Araújo et al., 2011; 2019). These correlative models relate the current distribution of species to

environmental variables, with the aid of algorithms, defining its climatic niche in environmental space, and thus projecting this niche model to geographic space and mapping the suitable areas for the species (Peterson et al., 2011; Bellard et al., 2012). This projection can be even transferred to past or future climate change scenarios (Peterson et al., 2011; Bellard et al., 2012). Although ENM applications in climate change are widespread for terrestrial organisms, studies on the vulnerability of biodiversity to climate change are particularly scarce for marine organisms (Robinson, 2011; 2017; Melo-Merino et al., 2020). In the rank of application issues, literature reviews about ENM in marine realm positioned climate change in the fourth (Robinson et al., 2011), third (Robinson et al., 2017), and second position (Melo-Merino et al., 2020). Although increasing, applications to evaluate impacts of climate change represent only 19% of the ENM studies for marine species (Melo-Merino et al., 2020). There is also a very clear geographical bias involving studies using ENM in marine realm, with most applications concerned the North Atlantic Ocean, often along the northeastern coast of the United States of America and the coast of Europe (Robinson et al., 2017; Melo-Merino et al., 2020).

For future projections, global predictions across hundreds to thousands of marine species indicated a general trend of range expansion and shift to higher latitudes (Cheung et al., 2009; Molinos et al., 2015). However, there is a certain divergence in these regards. In some recent and more restricted studies, for example, range contractions are predicted instead of expansions (e.g. Durante et al., 2017; Wabnitz et al., 2018; Zhang et al., 2019). In fact, for some temperate seaweeds, Juterbock (2013) found a possible displacement of areas suitable for higher latitudes in climate change scenarios, even leading to habitat loss at latitudes less than 45°. This same shift trend for higher latitudes is predicted for a species of shrimp in Antarctica (Basher, 2016) and for the benthic macrofauna of the English Channel (Rombouts, 2012). Some tuna species - the temperate tunas (albacore, Atlantic bluefin, and southern bluefin) and the tropical bigeye tuna - are expected to shrink in the tropics and undergo a poleward shift (Erasquin-Extramiana, 2019). For Japanese whiting *Sillago japonica* Zhang et al. (2019) pointed to a shift to the north of their distribution, towards higher latitudes. However, in a literature review on the use of ecological niche models applied to seaweed invasions, Marcelino (2015) reported that several studies involving species in this group have expanded their distribution to lower latitudes (e.g. near Ecuador). For skipjack and

yellowfin tunas, Erauskin-Extramiana (2019) have shown that they will be more abundant in tropical waters even with global warming.

Here, we evaluated the possible effects of future climate change on *Epinephelus itajara*, an iconic threatened species from tropical Atlantic (Craig et al., 2009; Bertoni et al., 2018). Goliath grouper, as it is usually called, is the largest grouper in the Atlantic Ocean, which can exceed 200kg and 2m. Its historical distribution is thought to extend from the southern Brazil to Florida, in the Western Atlantic, and from Congo to Senegal, in the Eastern Atlantic Ocean. Juveniles of this species are found in estuarine environments such as puddles, mangroves and seagrass. When individuals reach a certain size, an ontogenetic shift occurs in the habitat use towards estuaries and natural and artificial coral reefs. Due to its curiosity and fearlessness in the face of divers, the Goliath grouper suffered a lot from overfishing. Until recently, the species was categorized as Critically Endangered by the IUCN Red List. In 2018 it was recategorized as Vulnerable due to conservation efforts, like fishing prohibitions in United States and Brazil. In the African coast there are few *bonafide* records of the species in the last decades (see Chapter 1). The Goliath grouper is considered a tropical species, living generally in temperatures above 14° C (Gilmore et al., 1978). Therefore, with a possible tropicalization of the temperate marine areas, would also be possible an expansion and shift of *E. itajara* geographic distribution towards higher latitudes. In addition, a climate change study from LGM to the present Minsky (2017) predicted an increase in the suitable area for the Goliath grouper following the past increase in global temperature in the Middle Holocene.

In the present study we investigated whether the same trend predicted under past global warming would be repeated in a future climate change scenario in 2100. We hypothesized that the species will be affected by global warming in a future climate change scenario through (i) an expansion and (ii) a distribution shift of suitable areas towards higher latitudes. To test our hypotheses, we applied a correlative approach based on ecological niche modeling with the addition of habitat predictors, which was developed in the Chapter 2 and demonstrated to improve model accuracy and predictions. We also assessed the level of protection coverage of the future suitable areas predicted for 2100 and of the stable areas for *E. itajara* under the current Marine Protected Areas network.

3.2 MATERIALS AND METHODS

In the previous chapter, we demonstrated that the use of environmental variables that represent habitats used by the species can substantially improve the results of ecological niche modeling, with implications for biogeography and conservation. Therefore, following those results, in this chapter we used the EnvHabs model (see details in Chapter 2) to assess the exposure of suitable areas for *Epinephelus itajara* to the climate change projected for 2100. Considering that generation length of the species is 21.5 years (Bertoncini et al., 2018), we choose the year 2100 following the IUCN guidelines that sets a period of three generations for change analysis in risk assessments (IUCN, 2019).

3.2.1 Environmental and habitat variables for the RCP 8.5 scenario

We obtained the projections of the variables Salinity.Lt.min, Temperature.Lt.min for the year 2100, under the RCP 8.5 scenario, from the Bio-ORACLE v2.0 database (Assis et al., 2017). These future layers were produced for 2100 by averaging data from distinct AOGCMs provided by the CMIP 5. The representative concentration pathway RCP 8.5 is a "business as usual" scenario, which assumes high population growth, slow economic growth, resulting in high emissions of greenhouse gases and demand for energy, combined with few changes in climate change policies (reviewed by Moss et al., 2010).

The distance variables for the habitats were created through the projection of the ecological niche models of the respective habitats for the future climate change scenario, using the same procedures detailed in Appendix 6.

3.2.2 Ecological niche modeling

To generate ENMs we used the five algorithms (GLM, BRT, Maxent, SVM, and RBF) and the same parameterizations applied in the Chapter 2 (see details in section 2.2.3). We built consensus ENMs using only good and very good models ($TSS \geq 0.70$), and applying TSS weighted average ensemble method (see details in section 2.2.4). To provide a spatial estimate of the uncertainty, we used standard deviation of those outputs to create a uncertainty map. We also classify the suitability values of the ensemble model as Moderate Suitability (MS) and High Suitability (HS), using the suitability intervals generated by the Boyce Index (see details in section 2.2.4).

3.2.3 Exposure and impacts of climate changes in the RCP 8.5 scenario for 2100

The analysis of the exposure of suitable areas for *Epinephelus itajara* and impacts associated to future climate changes (RCP 8.5 scenario for 2100) was performed in a pixel-by-pixel basis (Hu et al. 2010, Hu & Jiang 2011). Following the methodology proposed by Thuiller et al. (2005) and using map algebra in ArcMap 10.5, we calculated the pixel loss in the model (pixels that are suitable in the present scenario, but are not suitable in the future), pixel gain (pixels that are not suitable in the present, but become suitable in the future), and stable pixels (pixel suitable in both time slices), for pixels classified as Highly Suitable (HS) and as Moderately Suitable (MS). Using zonal statistics, we calculated the extensions and percentages of the Moderately Suitable (MS) and Highly Suitable (HS) areas and the stable areas for the species in the future network of Marine Protected Areas. Data on MPAs (IUCN categories I-VI) were obtained from the Worldwide Database on Protected Areas (WDPA, <http://protectedplanet.net>). Clipping procedures, spatial overlap, map algebra, zonal statistics and area calculations were performed in ArcMap10.5. To avoid distortions in area calculations the data were converted to the World Equidistant Cylindrical Projection.

3.3 RESULTS

The results of the ecological niche modeling for the Goliath grouper in 2100 indicated that suitable areas predicted for this future climate change scenario sum 9×10^6 km², with 4×10^6 km² of highly suitable areas and 5×10^6 km² of moderately suitable areas. This represents a potential expansion of 164.7% compared to the current suitable areas, which totalize 3.4×10^6 km² (MS = 2.0×10^6 km², HS = 1.4×10^6 km²). If we consider only the pixel gain, i.e. pixels that were not suitable in the present, but become in the future, the area gain sums 8×10^6 km² (HS = 3.8×10^6 km², MS = 4.1×10^6 km²). The stable areas, i.e. pixels identified as suitable in the present and in the future, sum only 1×10^6 km² (HS = 1.9×10^5 km², MS = 8.1×10^5 km²), representing 29.4% of the suitable areas in the present and 11.1% of the suitable areas in the future.

The distribution of suitable areas in the future (Figures 3.2, 3.3) had a different pattern than the observed for the suitable areas in the present (Figure 3.1), described in

the Chapter 2 (section 2.3.3 of the Results). In general, there were latitudinal shifts of the suitable areas towards the poles predicted for 2100 (Figure 3.2). Higher concentrations of suitable areas in the present are observed in the tropical regions, while in the future are observed in subtropical and temperate regions. Considering only HS pixels, the distribution of the future suitable areas is located mainly below the Tropic of Capricorn and above the Tropic of Cancer. When considering only MS pixels, the distribution of the future suitable areas extends along almost all the African and American coast, but with some gaps. The distribution of stable areas for HS pixels occurs mainly in the regions of Florida coast, the mouth of the Amazon River, and in the southeastern coast of Brazil (Figure 3.2). For MS pixels, the distribution of the stable areas extends along almost all the African coast, Gulf of Mexico and the Caribbean islands of Central America.

The standard deviation map derived from our predictions showed that the uncertainties are generally low along the coastal zone and major part of the study area. Higher values are restricted to peripheral regions like north and south extreme limits or open water surrounding some islands further from the coast (Figure 3.3).

Likewise in the present, with only 6.4% of suitable areas protected, the results of future ENM indicated that a low percentage of predicted suitable areas for *Epinephelus itajara* in 2100 is under some kind of protection by the current MPAs network. Only 9.3% of the areas of HS is under some protection, while only 0.3% of the areas of MS is under protection. Considering only the stable areas, 19% of the HS pixels and 0.9% of the MS pixels is under some protection by the current MPAs network. Furthermore, the higher concentrations of the suitable areas in some protection predicted for 2100 are located in Tristan Gough and Azores Canaries Madeira ecoregions.



Figure 3.1: Environmental suitability for *Epinephelus itajara* predicted by the ensemble model in the present. Dark blue = High Suitability (HS), Light blue = Moderate Suitability (MS).

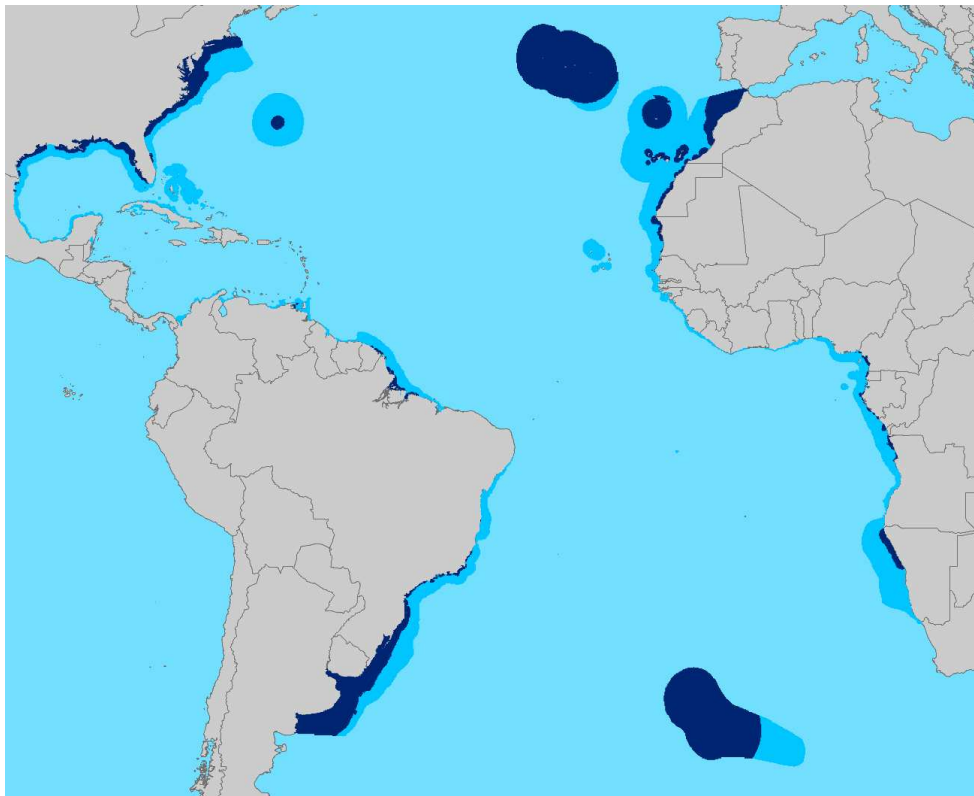


Figure 3.2: Environmental suitability for *Epinephelus itajara* predicted by the ensemble model in the future. Dark blue = High Suitability (HS), Light blue = Moderate Suitability (MS).



Figure 3.3: Distribution of the stable areas for *Epinephelus itajara*. Green area represent stable areas for High Suitability (HS) pixels and yellow areas represent stable area for Moderate Suitability (MS) pixels.

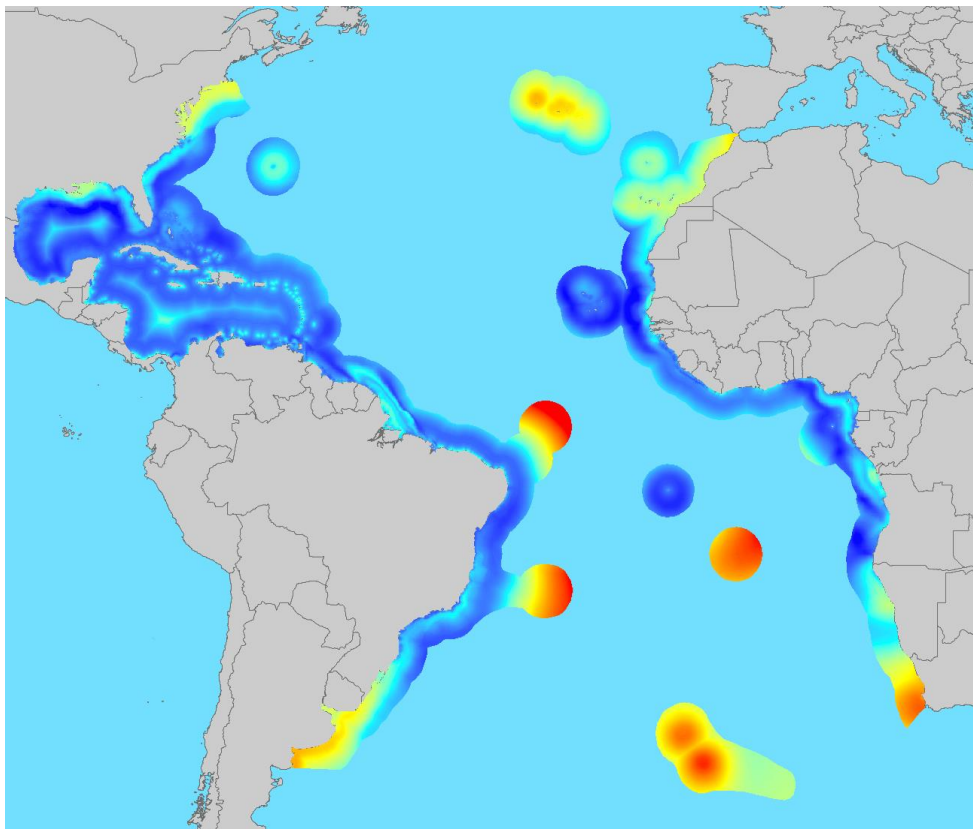


Figure 3.3: Plot of the continuous representation of the standard deviation for all models. Bluish light colors represent low deviation and reddish warm colors represent high deviation.

3.4 DISCUSSION

In this study we applied an ecological niche modeling approach, with the inclusion of habitat predictors, to test hypotheses about the effects of future climate change on the geographic distribution of *Epinephelus itajara*. The application of the EnvHabs niche model allowed to evaluate climate change impacts for the target species taking into account the synergistic effect of changes in temperature and salinity, together with the change that its habitats will undergo in the future scenario.

As we expected, the predicted suitable areas for the Goliath grouper in 2100 showed a potential expansion (1.5-fold) compared to the current distribution. This result agrees with studies that indicated a general trend of range expansion for marine species in future climate change scenarios (Cheung et al., 2009; Molinos et al., 2015; Morley et al., 2018), and disagree with recent studies that predicted range contractions instead of expansions (Durante et al., 2017; Wabnitz et al., 2018; Zhang et al., 2019). Likewise, according to our expectations, the distribution of suitable areas for *E. itajara* predicted for 2100 presented a poleward latitudinal shift, with higher concentrations of suitable areas located in subtropical and temperate regions, instead of in the tropical regions as observed in the current distribution. This poleward shift followed the general trend reported in previous studies predicting changes in marine species distribution (Cheung et al., 2009; Rombouts, 2012; Hazen et al., 2013; Juterbock, 2013; Robinson et al., 2015; Molinos et al., 2015; Basher, 2016; Kleisner et al., 2017; Morley et al., 2018; Zhang et al., 2019), and disagree with some studies that predicted shifts toward lower latitudes (e.g. Marcelino, 2015; Erauskin-Extramiana, 2019).

Prediction uncertainty showed to be low along the study area, except in extreme peripheral regions like north and south limits or open waters around some islands further from the coast. Therefore, the higher concentrations of the suitable areas predicted for 2100 in such regions, like Tristan Gough and Azores Canaries Madeira ecoregions (Figure 3.1), should be considered with care. ENM hindcast and forecast are always associated with some uncertainties. Despite the long-standing and important role of ENMs to assess the impact of past and future global change on biodiversity, these correlative approaches present some shortcomings (Araújo & Guisan, 2006; Bellard et al. 2012; Peterson et al., 2018). One such problem is limited transferability of parameterised models to other times beyond the range of data used for model fitting, i.e. model extrapolation, that bring uncertainties to outputs. Modelling methods and

parameterization, together with climate global circulation models and scenarios, are considered the main influences to transferability and uncertainties in ENMs (Diniz-Filho et al., 2009; Buisson et al. 2010; Loyola et al., 2012; Thuiller et al., 2019). Ensemble modelling and consensus methods have been proposed as an alternative to reduce uncertainty across individual ENM methods and climate models (Araújo & New, 2007; Diniz-Filho et al., 2009; Marmion et al., 2009; Buisson et al. 2010; Thuiller et al., 2019). In addition, the inclusion of additional non-climate predictors can help improve the performance of ENMs and, in turn, their transferability (Regos et al., 2019). Therefore, we expect to have minimized uncertainties associated to our predictions through the application of a modelling approach involving ensemble forecasting with consensus of different climate models and ENM methods, and of the EnvHabs niche model, optimized with a set of additional habitat predictors (see Chapter 2).

According to our predictions for expansion and shift in suitable areas for the Goliath grouper along with the possible tropicalization of the temperate marine areas, we could expect that this tropical species can expand its geographic distribution to larger latitudes, even without knowledge about the upper limit of its thermotolerance. However, even with new suitable areas available in the future, like predicted here for *Epinephelus itajara*, for a species to advance poleward in response to climate change, populations must colonize such new regions. A successful colonization requires a species to disperse to a new location and maintain positive growth through either self-persistence or ongoing immigration (Bridle & Vines 2007). To undergo range expansion, species must first arrive a new region, which can depend on movement performance, derived from larval dispersal and/or adult mobility capacity.

In general, larvae represent the life stage with the high dispersal potential (Feary, 2013) and, although there is a high level of larval retention in the area where they were settled (40 - 60%) (Thorrold et al, 2001; Almany et al., 2007a; Harrison et al., 2012), the corollary effect of this is that about 40 - 60% of the larvae will be exported away to other coral reefs or estuarine regions (Leis et al. 2011) and ocean currents play an important role in this event (Leis et al 2011). However, with climate change driven by anthropogenic actions, there may be a weakening of these currents (Rahmstorf 2006, Liu et al 2017). On the other hand, for many large marine fishes, adult mobility serves as a better predictor of the rate of range expansion than larval dispersal strategy (Sunday et al. 2015).

But even when species can arrive in a new region, it does not guarantee establishment and subsequent positive population growth (Burgess et al. 2012, Sadowski et al. 2018). The success of a tropical species in colonizing a new environment (usually subtropical or temperate) depends on extrinsic factors (e.g. ocean currents) and intrinsic factors (e.g. environmental and physiological restrictions, abundance and latitudinal distribution, life history traits associated with nomadism and association with habitats) (Feary, 2013).

As fish need temperature for their bioenergetic processes (Fry, 1967; Hazel & Prosser, 1974; Houde, 1989; Clarke & Johnston, 1999), the thermal limits for minimum environmental temperature will dictate the success of tropical species in colonizing subtropical or temperate environments (Attrill & Power, 2002; Dulvy et al. 2008; Poertner & Farrell, 2008). Considering that non-tropical environments vary greatly in temperature throughout the year, the optimal thermal for key physiological processes will only be possible at specific times of the year (e.g. hot summer months) (Feary, 2013).

The success in dispersing to non-tropical environments can be associated to species biogeographic traits. There is a positive relationship between population density and geographic extent (Lawton, 1999; Roughgarden, 2009) and between vagrancy and latitudinal extent (Feary, 2013). However, the success of a tropical fish species migrating to a subtropical or temperate region also depends on how the populations are distributed, because even species with high density and wide range, but with a truncated or sparsely geographic distribution, may not be able to maintain viable populations since the source population will be very distant or will not be enough to maintain its sink population (Sorte et al., 2010).

Traits of life history also affect the success in dispersing to non-tropical environments. There is a positive relationship between fish body size and the number of gametes produced (Weatherly, 1972; Thresher, 1984), which can substantially increase the likelihood of larval dispersion over long distances (Law, 1993; Munday & Jones, 1998). In fact, Feary (2013) demonstrated that fish with larger body size are more likely to have assemblages of expatriates compared to fish of smaller size in a possible climate change environment. In addition, life traits as the type of parental care for offspring (Thresher 1984; Brogan 1994; Lo-Yat et al., 2006), swimming ability in the larval phase (Stobutzki & Bellwood, 1994; 1997; Leis et al., 1996; Leis & Carson-Ewart 1997, 2003; Fisher, 2005), body size in the settlement phase (Sponaugle & Grorud-Covert, 2006;

Sponaugle et al., 2011), and dependence on specific habitats to settle (Feary 2013), are other factors that can facilitate or not the dispersion of these species. Indeed, Feary (2013) demonstrated that, in a scenario of climate changes, low parental care, high swimming capacity, larger body size in the settlement phase and larvae, and habitat generalist larvae are the main traits associated to the success in colonizing non-tropical environments.

Along with suitable environmental conditions, to persist in a new region species need to find food resources. Accordingly, species that successfully colonize new habitats tend to be ecological generalists that have flexibility in habitat and diet, such as omnivores as opposed to herbivores or predators (Sunday et al., 2015).

Although with the potential to expand the range polewards, *Epinephelus itajara* might find challenges to migrate into new areas due to some life history and physiological traits. The Goliath grouper is a site specific species with low activity (Craig, 2015), predator food habits (Artero et al., 2015a), and reproduction through seasonal aggregations, which involve many individuals migrating to specific sites to spawn (Koenig et al., 2016). Large juveniles and adults can have difficulties to successfully migrate to novel areas due to their strong site fidelity and low activity (Koenig et al., 2007; Bertoncini et al., 2018). Young juveniles are known to strongly depend on mangroves, and in a minor proportion on coral reefs and rocky tide pools (Koenig et al., 2007; Lobato et al., 2016), therefore, juveniles can face some challenges to occupy novel areas, especially if they are unable to settle in these habitats. Furthermore, to sustain viable populations on these novel areas the Goliath grouper will strongly depend on tropical source populations, since the individuals that will live in these novel regions or at the border of the distribution will be spending much of the energetic reserve to deal with this novel environment (Feary et al., 2013; Phillips and Pérez-Ramírez, 2018). It is worth mentioning that much of the Brazilian northeastern coastal region is predicted to be no longer suitable in 2100. This is particularly worrying because this gap in the distribution of suitable areas can interfere with the genetic flow between the populations of Goliath grouper, which has already been suffering from population genetic structuring (Craig et al., 2009; Damasceno et al., 2015).

More important than the distribution of suitable areas in the future is the distribution of environmentally stable areas. Climatic stable areas are important to providing suitable habitats over time (Carnaval et al., 2009) and maintaining genetic diversity (Assis et al., 2014; Assis et al., 2016; Carvalho et al., 2017), serving as refuges

and increasing the possibility of long-term persistence of species. Stable areas favor evolutionary process (Werneck et al., 2012), being regions associated with high endemism, species diversity and intraspecific diversity (Graham et al., 2006; Hewitt, 2004; Werneck et al., 2012). Our results indicated that part of the mouth of the Amazon River, part of the southern region and the state of Florida can represent stable regions, when considering only the HS pixels. Currently these regions are extremely important because they are known to harbor seasonal aggregations and nurseries for the species (Giglio et al., 2014b; Zapelini et al., 2017). These HS stable areas could serve as a refuge for the species in a climate change scenario.

Regarding the temporal dynamics, as expected, from the present to the future we found a similar pattern of suitable areas expansion predicted under past global warming, although the trend for potential range expansion in 2100 (~165%) had been higher than that from LGM to Middle Holocene (~60%) (Minsky, 2017). However, past and future dynamics present some important differences. The suitable areas expansion predicted in the past global warming did not show distributional changes in latitude, a distinct situation from that predicted for the future with a poleward shift. From LGM to the present, the stable areas (i.e., areas that remain suitable along time) represented more than 80% of the current suitable areas (Minsky, 2017). Differently from the past, the distribution of stable areas predicted in the future corresponded to less than 30% of the current suitable areas. Therefore, the restricted distribution of stable areas in 2100 raises a concern for the future persistence of *E. itajara*.

In the same direction, the current MPAs network seems to offer a limited coverage for the target species both in the present and in the future, with less than 10% of the suitable areas and to less than 20% of the stable areas under some level of protection. Considering that the higher concentrations of the suitable areas under MPAs are located in Tristan Gough and Azores Canaries Madeira ecoregions, where predictions have higher uncertainty, the protected status of potential habitats for the Goliath grouper in the future can be even worse.

In summary, our findings suggest that future climate change may have a positive impact on the potential range extent of *Epinephelus itajara*, promoting a gain of suitable areas for this species. This could reduce the species' vulnerability to future climate changes. However, the distribution of future suitable areas will undergo a poleward shift. In a possible negative outcome of such change, areas currently suitable and occupied by the Goliath grouper could become unsuitable and suffer local extirpations

in the future. Besides, new suitable areas predicted for the future can be prevented to be occupied by *E. itajara* due to dispersion and biotic (e.g. prey, competitors, pathogens) restrictions. It is important to consider both biotic and abiotic drivers in facilitating or limiting range expansions, though the relative importance of these processes remains poorly understood (Louthan et al. 2015). Studies on the ecology and behavior of Goliath grouper larvae are necessary to better predict the success of colonizing new regions. Studies on Goliath grouper thermotolerance are needed to better assess the climate change exposure. Evidences on a congeneric species demonstrated that spawning individuals present narrow thermal tolerance range, indicating that this life stage may be a bottleneck constraining response to climate change (Asch & Erisman, 2018). Finally, we emphasize the need to expand MPAs network to increase the coverage of protection on suitable areas for *Epinephelus itajara*, especially in regions with stable areas, to contribute to the long-term persistence of this vulnerable iconic species in the future.

General Conclusions

This master thesis highlights the importance of applying procedures of filtering and selection of the occurrence data and predictor variables to improve ENMs. Using the Goliath grouper (*Epinephelus itajara*) as a model species we first assess the effect of different datasets of occurrence data on ENMs performance and outputs, then the effect of different datasets of environmental predictors and, at last, we applied the best approaches to develop an improved analysis of future climate change impacts on the range of *E. itajara*.

In the first chapter, we developed a framework that was able to support decision about the use of unreliable records in ecological niche modeling for the target species, shedding light on how uncertainty in occurrence data can affect ENMs and overall measures of model accuracy. The main conclusions from this chapter were that the use of uncertain occurrence records decrease the models' performance, increasing their omission error and decreasing their ability to project the models from the environmental space to the geographical space, leading to low power to predict suitable areas. But we highlighted that the decision about the exclusion of these uncertain and/or biased records should be take with care, because if the records with low uncertainty are not enough to represent the entire environmental niche of the species such an exclusion can be unfeasible. We also called the attention for the high proportion of records with high uncertainty and bias among data available from online occurrence repositories, which point to the need for data cleaning and filtering in biodiversity studies, and also to the need to improve and expand the occurrence records databases.

In the second chapter, we compared models developed with different combinations of environmental predictors to assess the effect of adding habitat variables on ENMs performance. Our results suggest that habitat predictors in combination with climate have a strong influence on ENMs accuracy and suitability predictions. The model that combined climate/salinity and habitat predictors for different habitats (EnvHabs) showed the best performance according to all analysed metrics (Sensitivity, Specificity, TSS and Boyce Index). Furthermore, models that include habitat variables (EnvHabs and HabOnly) seems to be more able to predict the biogeographic patterns in more detail and the historical distribution of the species. We also highlight the need to

choose carefully the predictors to calibrate models, and suggest that add non-climate variables to ENMs, particularly habitat predictors, should be recommended. In addition, our results about exposure to anthropogenic stressors for suitable areas predicted by both best models (EnvHabs and HabOnly) revealed a worrying situation for *E. itajara*. The species range are under great exposure, where most of suitable area extent (>95%) face some type of human pressure, with similar values of impact for all antropogenic stressors analyzed (artisanal fishing; demersal, non-destructive, lowbycatch fishing; ocean pollution and population pressure). In the same direction, the low percentage (>9,2%) of predicted suitable areas under some kind of protection by the current MPAs network increasing the concern about the target species persistence.

Finally, in the third chapter, we applied the best ENM approach identified in the second chapter to improve the assessment of future climate change impacts on the geographic distribution of *Epinephelus itajara*. Using the EnvHabs model, improved by the inclusion of habitat predictors, we were able to assess climate change impacts in 2100 scenario taking into account the synergistic effect of changes in temperature and salinity, along with the change that species habitats will undergo in the future. Our findings suggest that future climate change may have a positive impact on the potential range extent of *Epinephelus itajara*, promoting a gain (1.5-fold) in the suitable areas compared to the current distribution. Although this could reduce the species' vulnerability to future climate changes, the distribution of suitable areas in 2100 will undergo a poleward shift. We highlighted that this change can result in a negative impact if areas currently suitable and occupied by the Goliath grouper turn unsuitable and undergo local extirpations in the future. Furthermore, new suitable areas predicted for the future can be prevented to be occupied by *E. itajara* due to dispersion and biotic (e.g. prey, competidors, pathogens) restrictions. We identified stable areas that could serve as a refuge for the species in a climate change scenario, located in the mouth of the Amazon River, in part of the southeastern and southern Brazilian coast and the state of Florida. Differently from the past global warming (LGM to Middle Holocene), when 80% of the suitable areas were maintained as stable areas, the extent of stable areas predicted in the future corresponded to less than 30% of the current suitable areas. Therefore, the restricted distribution of stable areas in 2100 raises a concern for the future persistence of *E. itajara*. We also found a worrying situation regarding habitat protection, indicating that the current MPAs network offers a limited coverage for the target species both in the present and in the future, with less than 10% of the suitable

areas and less than 20% of the stable areas under some level of protection. Therefore, an expansion of the MPAs network, especially in regions with stable areas for *E. itajara*, it is highly recommended to contribute to the long-term persistence of this vulnerable iconic species in the future.

References

- Abecasis, D., Afonso, P., Erzini, K., 2014. Combining multispecies home range and distribution models aids assessment of MPA effectiveness. *Mar. Ecol. Prog. Ser.* 513, 155–169. <https://doi.org/10.3354/meps10987>
- Ahrends, A., Burgess, N.D., Gereau, R.E., Marchant, R., Bulling, M.T., Lovett, J.C., Platts, P.J., Wilkins Kindemba, V., Owen, N., Fanning, E., Rahbek, C., 2011. Funding begets biodiversity. *Divers. Distrib.* 17, 191–200. <https://doi.org/10.1111/j.1472-4642.2010.00737.x>
- Alexander Pyron, R., Burbrink, F.T., 2009. Lineage diversification in a widespread species: Roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Mol. Ecol.* 18, 3443–3457. <https://doi.org/10.1111/j.1365-294X.2009.04292.x>
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Almany, G.R., Berumen, M.L., Thorrold, S.R., Planes, S., Jones, G.P., 2007. Local replenishment of coral reef fish populations in a marine reserve. *Science* (80-.). 316, 742–744. <https://doi.org/10.1126/science.1140597>
- Alt, K.G., Kochmann, J., Klimpel, S., Cunze, S., 2019. Improving species distribution models of zoonotic marine parasites. *Sci. Rep.* 9, 1–10. <https://doi.org/10.1038/s41598-019-46127-6>
- Amano, T., Sutherland, W.J., 2013. Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proc. R. Soc. B Biol. Sci.* 280. <https://doi.org/10.1098/rspb.2012.2649>
- Amboni, M.P.M., 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. <https://doi.org/10.1080/13658816.2012.717627>
- Anderson, R. P. 2012. Harnessing the world’s biodiversity data: promise and peril in ecological niche modeling of species distributions. *Ann. N. Y. Acad. Sci.* 1260: 66–80.
- Araújo, M. B. & Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688.
- Araújo, M.B., Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16: 743–753.
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>

- Araújo, M.B., Peterson, A.T. 2012. Uses and misuses of bioclimatic envelope modelling. *Ecology*, 93: 1527-1539.
- Araújo, M.B.M.B., Garcia, R.A.R.A., Naimi, B., Rahbek, C., Barbosa, A.M.M., Anderson, R.P.R.P., Beale, C.M.C.M., Dormann, C.F.C.F., Early, R., Guisan, A., Maiorano, L., O'Hara, R.B., Zimmermann, N.E.N.E., Barbosa, A.M.M., Beale, C.M.C.M., Dormann, C.F.C.F., Early, R., Garcia, R.A.R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E.N.E., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5, 1–12. <https://doi.org/10.1126/sciadv.aat4858>
- Artero, C., Koenig, C.C., Richard, P., Berzins, R., Guillou, G., Bouchon, C., Lampert, L., 2015a. Ontogenetic dietary and habitat shifts in goliath grouper *Epinephelus itajara* from French Guiana. *Endanger. Species Res.* 27, 155–168. <https://doi.org/10.3354/esr00661>
- Artero, C., Murie, D.J., Koenig, C.C., Berzins, R., Bouchon, C., Lampert, L., 2015b. Age, growth, and mortality of the Atlantic goliath grouper *Epinephelus itajara* in French Guiana. *Endanger. Species Res.* 28, 275–287. <https://doi.org/10.3354/esr00691>
- Asch, R.G., Erisman, B. 2018. Spawning aggregations act as a bottleneck influencing climate change impacts on a critically endangered reef fish. *Diversity and Distributions*, 24:1712–1728.
- Assis J, Coelho NC, Lamy T, Valero M, Alberto F, Serrão EA. 2016. Deep reefs are climatic refugia for genetic diversity of marine forests. *Journal of Biogeography*, 43:833–844.
- Assis, J., Serrão, E.A., Claro, B., Perrin, C. & Pearson, G.A. 2014. Climate-driven range shifts explain the distribution of extant gene pools and predict future loss of unique lineages in a marine brown alga. *Molecular Ecology*, 23, 2797–810.
- Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E.A., De Clerck, O., 2018. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* 27, 277–284. <https://doi.org/10.1111/geb.12693>
- Assis, J., Tyberghein, L., Bosh, S., Verbruggen, H., Serrao, E. A., & De Clerck, O. 2017. Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27: 277-284.
- Attrill, M.J., Power, M., 2002. Climatic influence on a marine fish assemblage. *Nature* 417, 275–278. <https://doi.org/10.1038/417275a>
- Aubry KB, Raley CM, McKelvey KS. 2017. The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS ONE* 12(6): e0179152.
- Austin, M., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Modell.* 157, 101–118. <https://doi.org/10.1007/BF02205449>

- Austin, M.P., Meyers, J.A., 1996. Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. *For. Ecol. Manage.* 85, 95–106. [https://doi.org/10.1016/S0378-1127\(96\)03753-X](https://doi.org/10.1016/S0378-1127(96)03753-X)
- Austin, M.P., Van Niel, K.P. 2011. Improving species distribution models for climate change studies: variable selection and scale. *J. Biogeography*, 38: 1-8.
- Báez, J.C., Barbosa, A.M., Pascual, P., Ramos, M.L., Abascal, F., 2020. Ensemble modeling of the potential distribution of the whale shark in the Atlantic Ocean. *Ecol. Evol.* 10, 175–184. <https://doi.org/10.1002/ece3.5884>
- Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P., Beck, J., 2013. Mapping the biodiversity of tropical insects: Species richness and inventory completeness of African sphingid moths. *Glob. Ecol. Biogeogr.* 22, 586–595. <https://doi.org/10.1111/geb.12039>
- Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3: 327-338.
- Barnett, T. P., Pierce, D. W., AchutaRao, K. M., Gleckler, P. J., Santer, B. D., Gregory, J. M., Washington, W. M. 2005. Penetration of human-induced warming into the world's oceans. *Science*, 309(5732): 284-287.
- Barnett, T. P., Pierce, D. W., Schnur, R. 2001. Detection of anthropogenic climate change in the world's oceans. *Science*, 292: 270–274.
- 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. Modell.* 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Basher, Z., Costello, M.J., 2016. The past , present and future distribution of a deep-sea shrimp in the Southern Ocean. <https://doi.org/10.7717/peerj.1713>
- Baumann, H., Talmage, S.C., Gobler, C.J., 2012. Reduced early life growth and survival in a fish in direct response to increased carbon dioxide. *Nat. Clim. Chang.* 2, 38–41. <https://doi.org/10.1038/nclimate1291>
- Beck, M.W., 1995. Size-Specific Shelter Limitation in Stone Crabs: A Test of The Demographic Bottleneck Hypothesis 76, 968–980.
- Beck, M.W., Heck, K.L., Able, K.W., Childers, D.L., Eggleston, D.B., Gillanders, B.M., Halpern, B., Hays, C.G., Hoshino, K., Minello, T.J., Orth, R.J., Sheridan, P.F., Weinstein, M.P., 2001. The Identification, Conservation, and Management of Estuarine and Marine Nurseries for Fish and Invertebrates. *Bioscience* 51, 633. [https://doi.org/10.1641/0006-3568\(2001\)051\[0633:TICAMO\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0633:TICAMO]2.0.CO;2)
- Beger, M., Jones, G.P., Munday, P.L., 2003. Conservation of coral reef biodiversity: A comparison of reserve selection procedures for corals and fishes. *Biol. Conserv.* 111, 53–62. [https://doi.org/10.1016/S0006-3207\(02\)00249-5](https://doi.org/10.1016/S0006-3207(02)00249-5)

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., Courchamp, F., 2012. Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15, 365–377. <https://doi.org/10.1111/j.1461-0248.2011.01736.x>

Benevides, E.A., Vallinoto, M.N.S., Fetter Filho, A.F.H., de Souza, J.R.B., Silva-Oliveira, G., Freitas, M.O., Ferreira, B.P., Hostim-Silva, M., Bertoncini, A.A., Blanchard, F., Torres, R.A., 2014. When physical oceanography meets population genetics: The case study of the genetic/evolutionary discontinuity in the endangered goliath grouper (*Epinephelus itajara*; Perciformes: Epinephelidae) with comments on the conservation of the species. *Biochem. Syst. Ecol.* 56, 255–266. <https://doi.org/10.1016/j.bse.2014.06.004>

Benjamin S. Halpern, Melanie Frazier, Jamie Afflerbach, Julia S. Lowndes, Fiorenza Micheli, Casey O'Hara, Courtney Scarborough, Kimberly A. Selkoe. Recent pace of change in human impact on the world's ocean. *Scientific Reports*, 2019; 9 (1)

Bertoncini, A.A., Aguilar-Perera, A., Barreiros, J., Craig, M.T., Ferreira, B., Koenig, C. 2018. *Epinephelus itajara*. The IUCN Red List of Threatened Species 2018: e.T195409A46957794. <https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T195409A46957794.en>.

Bilbao, R. A. F., Gregory, J. M., Bouttes, N., Palmer, M. D., Stott, P. 2019. Attribution of ocean temperature change to anthropogenic and natural forcings using the temporal, vertical and geographical structure. *Clim. Dynam.*, 53: 5389–5413.

Bindoff, N.L. et al. 2013. Detection and Attribution of Climate Change: from Global to Regional. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F. et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-Qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol.* 8. <https://doi.org/10.1371/journal.pbio.1000385>

Boitani, L., Maiorano, L., Baisero, D., Falcucci, A., Visconti, P. & Rondinini, C. 2011. What spatial data do we need to develop global mammal conservation strategies? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366, 2623–2632.

Boitani, L., Sinibaldi, I., Corsi, F., De Biase, A., Carranza, I.D.I., Ravagli, M., Reggiani, G., Rondinini, C., Trapanese, P., 2008. Distribution of medium- to large-sized African mammals based on habitat suitability models. *Biodivers. Conserv.* 17, 605–621. <https://doi.org/10.1007/s10531-007-9285-0>

Booth, D.J., Feary, D., Kobayashi, D., Luiz, O., Nakamura, Y., 2017a. Tropical Marine Fishes and Fisheries and Climate Change. *Clim. Chang. Impacts Fish. Aquac.* 875–896. <https://doi.org/10.1002/9781119154051.ch26>

- Booth, D.J., Poloczanska, E., Donelson, J.M., Molinos, J.G., Burrows, M., 2017b. Biodiversity and Climate Change in the Oceans. *Clim. Chang. Impacts Fish. Aquac. I*, 63–89. <https://doi.org/10.1002/9781119154051.ch4>
- Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73–77.
- Bosch, S. et al. 2018. In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. *Diversity and Distributions*, 24: 144–157.
- Botkin, D.B., Saxe, H., Araújo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P., Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert, D.W., Loehle, C., Margules, C., New, M., Sobel, M.J., Stockwell, D.R.B., 2007. Forecasting the Effects of Global Warming on Biodiversity. *Bio Sci.* 57, 227–236. <https://doi.org/10.1641/B570306>
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A. 2002. Evaluating resource selection functions. *Ecological Modelling*, 157: 281–300.
- Boykoff, M. and O. Pearman, 2019. Now or Never: How Media Coverage of the IPCC Special Report on 1.5C Shaped Climate-Action Deadlines. *One Earth* 1 (3)
- Bradie, J., & Leung, B. (2016). A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography*, 44, 1344–1361.
- Bradshaw CJA, Ehrlich PR, Beattie A, Ceballos G, Crist E, Diamond J, Dirzo R, Ehrlich AH, Harte J, Harte ME, Pyke G, Raven PH, Ripple WJ, Saltré F, Turnbull C, Wackernagel M and Blumstein DT. 2021. Underestimating the Challenges of Avoiding a Ghastly Future. *Front. Conserv. Sci.* 1:615419.
- Branch, T.A., DeJoseph, B.M., Ray, L.J., Wagner, C.A., 2013. Impacts of ocean acidification on marine seafood. *Trends Ecol. Evol.* 28, 178–186. <https://doi.org/10.1016/j.tree.2012.10.001>
- Brant, S. V., Ortí, G., 2003. Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): Past fragmentation and postglacial recolonization. *Mol. Ecol.* 12, 1435–1449. <https://doi.org/10.1046/j.1365-294X.2003.01789.x>
- Brogan, M.W., 1994. Distribution and retention of larval fishes near reefs in the Gulf of California. *Mar. Ecol. Prog. Ser.* 115, 1–14. <https://doi.org/10.3354/meps115001>
- Brown J. H., Gibson, A. C. 1983. *Biogeography*. Mosby. Missouri: St. Louis.
- Brown, J.H., Stevens, G.C., Kaufman, D.M., 1996. The geographic range: Size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Syst.* 27, 597–623. <https://doi.org/10.1146/annurev.ecolsys.27.1.597>
- Brown, J.L., 2014. SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol.* 5, 694–700. <https://doi.org/10.1111/2041-210X.12200>

- Bucklin, D.N. et al. 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and Distributions*, 21: 23–35.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Glob. Chang. Biol.* 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>
- Bullock, M.F.G. y M.E.M.L.H.M.D.M., 1992. Age, growth, and reproduction of jawfish *Epinephelus itajara* in the eastern Gulf of Mexico. *Fish. Bull.*,.
- Byrne, M., Gonzalez-Bernat, M., Doo, S., Foo, S., Soars, N., Lamare, M., 2013. Effects of ocean warming and acidification on embryos and non-calcifying larvae of the invasive sea star *Patiriella regularis*. *Mar. Ecol. Prog. Ser.* 473, 235–246. <https://doi.org/10.3354/meps10058>
- Capinha, C., Rocha, J., Sousa, C. A. 2014. Macroclimate determines the global range limit of *Aedes aegypti*. *EcoHealth*, 11(3),420–428.
- Castellanos, A. A., Huntley, J. W., Voelker, G., Lawing, A. M. 2019. Environmental filtering improves ecological niche models across multiple scales. *Methods in Ecology and Evolution*,10: 481–492.
- Catucci, E., Scardi, M., 2020. A Machine Learning approach to the assessment of the vulnerability of *Posidonia oceanica* meadows. *Ecol. Indic.* 108, 105744. <https://doi.org/10.1016/j.ecolind.2019.105744>
- Cayuela, L., Golicher, D.J., Newton, A.C., Kolb, M., de Albuquerque, F.S., Arets, E.J.M.M., Alkemade, J.R.M., Pérez, A.M., 2009. Species Distribution Modeling in the Tropics: Problems, Potentialities, and the Role of Biological Data for Effective Species Conservation. *Trop. Conserv. Sci.* 2, 319–352. <https://doi.org/10.1177/194008290900200304>
- Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M., Silander, J.A., 2011. Point pattern modelling for degraded presence-only data over large regions. *J. R. Stat. Soc. Ser. C (Applied Stat.* 60, 757–776. <https://doi.org/10.1111/j.1467-9876.2011.01023.x>
- Chapman FA, García LN, Atencio VJ, Muñoz RJ, Silva A, Flores H. 2014. Low-salinity acclimation of juvenile marine goliath grouper *Epinephelus itajara*. *J Appl Aquacult* 26:179–186.
- Chatfield, B. S., Van Niel, K. P., Kendrick, G. A., Harvey, E. S. 2010. Combining environmental gradients to explain and predict the structure of demersal fish distributions. *Journal of Biogeography*, 37:593-605.
- Chen, D., and A. Getis. 1998. *Point Pattern Analysis (PPA)*. San Diego, CA: Department of Geography, San Diego State University
- Cheng L, Abraham J, Hausfather Z, Trenberth KE. 2019. How fast are the oceans warming? *Science*, 363:128–29.

- Cheng, H., Sinha, A., Cruz, F.W., Wang, X., Edwards, R.L., D’Horta, F.M., Ribas, C.C., Vuille, M., Stott, L.D., Auler, A.S., 2013. Climate change patterns in Amazonia and biodiversity. *Nat. Commun.* 4. <https://doi.org/10.1038/ncomms2415>
- Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R., Pauly, D., 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish Fish.* 10, 235–251. <https://doi.org/10.1111/j.1467-2979.2008.00315.x>
- Clarke, A., Johnston, N.M., 1999. Scaling of metabolic rate with body mass and temperature in teleost fish. *J. Anim. Ecol.* 68, 893–905. <https://doi.org/10.1046/j.1365-2656.1999.00337.x>
- Collins, A., 2014. An investigation into the habitat, behavior and opportunistic feeding strategies of the protected Goliath grouper (*Epinephelus itajara*) 171. [https://doi.org/USF Scholar Commons](https://doi.org/USF_Scholar_Commons)
- Collins, A.B., Barbieri, L.R., 2010. Behavior, Habitat and Abundance of the Goliath Grouper, *Epinephelus itajara*, in the Central Eastern Gulf of Mexico.
- Collins, A.B., Barbieri, L.R., McBride, R.S., McCoy, E.D., Motta, P.J., 2015. Reef relief and volume are predictors of Atlantic goliath grouper presence and abundance in the Eastern Gulf of Mexico. *Bull. Mar. Sci.* 91, 399–418. <https://doi.org/10.5343/bms.2015.1001>
- Cornelissen, B., Neumann, P., Schweiger, O., 2019. Global warming promotes biological invasion of a honey bee pest. *Glob. Chang. Biol.* 25, 3642–3655. <https://doi.org/10.1111/gcb.14791>
- Costa, H., Foody, G. M., Jiménez, S., Silva, L. 2015. Impacts of Species Misidentification on Species Distribution Modeling with Presence-Only Data. *International Journal of Geo-Information*, 4, 2496-2518.
- Craig, A., 2015. *Epinephelus itajara* , Atlantic Goliath Grouper 8235, 1–2.
- Craig, M.T., Graham, R.T., Torres, R.A., Hyde, J.R., Freitas, M.O., Ferreira, B.P., Hostim-Silva, M., Gerhardinger, L.C., Bertoni, A.A., Robertson, D.R., 2009. How many species of goliath grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species. *Endanger. Species Res.* 7, 167–174. <https://doi.org/10.3354/esr00117>
- Creed, J., Pires, D., Figueiredo, M., 2007. Biodiversidade marinha da Baía da Ilha Grande, Young.
- Dahlgren, C.P., Eggleston, D.B., 2000. Ecological processes underlying ontogenetic habitat shifts in a coral reef fish. *Ecology* 81, 2227–2240. [https://doi.org/10.1890/0012-9658\(2000\)081\[2227:EPUOHS\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[2227:EPUOHS]2.0.CO;2)
- Damasceno, J.S., Siccha-Ramirez, R., Morales, M.J.A., Oliveira, C., Torres, R.A., Costa, E.N., Silva-Oliveira, G.C., Vallinoto, M., Machado, L.F., Tosta, V.C., Farro, A.P.C., Hostim-Silva, M., 2015. Mitochondrial DNA evidences reflect an incipient population structure in Atlantic goliath grouper (*Epinephelus itajara*),

- Epinephelidae) in Brazil. *Sci. Mar.* 79, 419–429.
<https://doi.org/10.3989/scimar.04203.20A>
- Daniels, R.C., White, T.W., Chapman, K.K., 1993. Sea-level rise: Destruction of threatened and endangered species habitat in South Carolina. *Environ. Manage.* 17, 373–385. <https://doi.org/10.1007/BF02394680>
- Day, J.W., Templet, P.H., 1989. Consequences of sea level rise: Implications from the mississippi delta. *Coast. Manag.* 17, 241–257.
<https://doi.org/10.1080/08920758909362088>
- De Marco, P., Nóbrega, C.C., 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS One* 13.
<https://doi.org/10.1371/journal.pone.0202403>
- De Sylva, D.P., 1994. Distribution and ecology of ciguatera fish poisoning in Florida, with emphasis on the Florida Keys. *Bull. Mar. Sci.* 54, 944–954.
- Dennis, R.L.H., Thomas, C.D., 2000. Bias in butterfly distribution maps: The influence of hot spots and recorder's home range. *J. Insect Conserv.* 4, 73–77.
<https://doi.org/10.1023/A:1009690919835>
- Diniz-Filho, J.A.F., Mauricio Bini, L., Fernando Rangel, T., Loyola, R.D., Hof, C., Nogués-Bravo, D., Araújo, M.B., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography (Cop.)*. 32, 897–906. <https://doi.org/10.1111/j.1600-0587.2009.06196.x>
- Donelson, J.M., Munday, P.L., McCormick, M.I., Pitcher, C.R., 2012. Rapid transgenerational acclimation of a tropical reef fish to climate change. *Nat. Clim. Chang.* 2, 30–32. <https://doi.org/10.1038/nclimate1323>
- Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob. Ecol. Biogeogr.* 16, 129–138.
<https://doi.org/10.1111/j.1466-8238.2006.00279.x>
- Drake JM, Randin C, Guisan A. Modelling ecological niches with support vector machines. *J Appl Ecol.* Blackwell Publishing Ltd; 2006; 43: 424-432.
- Drinkwater, K.F., 2006. The regime shift of the 1920s and 1930s in the North Atlantic. *Prog. Oceanogr.* 68, 134–151. <https://doi.org/10.1016/j.pocean.2006.02.011>
- Dulvy, N.K., Rogers, S.I., Jennings, S., Stelzenmüller, V., Dye, S.R., Skjoldal, H.R., 2008. Climate change and deepening of the North Sea fish assemblage: A biotic indicator of warming seas. *J. Appl. Ecol.* 45, 1029–1039. <https://doi.org/10.1111/j.1365-2664.2008.01488.x>
- Durante, L. M., Cruz, I. C. S.; Lotufo, T. M.C. 2017. The effect of climate change on the distribution of a tropical zoanthid (*Palythoa caribaeorum*) and its ecological implications. *PeerJ* 6:e4777.
- Eklund A, Schull J. 2001. A stepwise approach to investigating the movement patterns and habitat utilization of goliath grouper, *Epinephalus itajara*, using conventional

tagging, acoustic telemetry and satellite tracking. *Electronic Tagging and Tracking in Marine Fisheries* 189–216.

El-Gabbas, A., Dormann, C.F., 2018a. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography (Cop.)*. 41, 1161–1172.

<https://doi.org/10.1111/ecog.03149>

El-Gabbas, A., Dormann, C.F., 2018b. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecol. Evol.* 8, 2196–2206. <https://doi.org/10.1002/ece3.3834>

Elith, J. & Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.

Ensing, D.J.; Moffat, C.E.; Pither, J. 2013. Taxonomic identification errors generate misleading ecological niche model predictions of an invasive hawkweed. *Botany* 91: 137–147.

Erauskin-Extramiana, M., Arrizabalaga, H., Hobday, A.J., Cabré, A., Ibaibarriaga, L., Arregui, I., Murua, H., Chust, G., 2019. Large-scale distribution of tuna species in a warming ocean. *Glob. Chang. Biol.* 25, 2043–2060. <https://doi.org/10.1111/gcb.14630>

ESRI. 2018. ArcMap 10.5. Environmental Systems Research Institute.

Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography (Cop.)*. 30, 609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>

Faith, D.P., Walker, P.A., 1996. Integrating conservation and development: Effective trade-offs between biodiversity and cost in the selection of protected areas. *Biodivers. Conserv.* 5, 431–446. <https://doi.org/10.1007/BF00056389>

Feary, D.A., Pratchett, M.S., J Emslie, M., Fowler, A.M., Figueira, W.F., Luiz, O.J., Nakamura, Y., Booth, D.J., 2013. Latitudinal shifts in coral reef fishes: Why some species do and others do not shift. *Fish Fish.* 15, 593–615. <https://doi.org/10.1111/faf.12036>

Feeley K. 2015. Are We Filling the Data Void? An Assessment of the Amount and Extent of Plant Collection Records and Census Data Available for Tropical South America. *PLoS ONE* 10(4): e0125629.

Feeley KJ, Silman MR. 2011. The data void in modeling current and future distributions of tropical species. *Global Change Biology* 17: 626–630.

- Fisher, R., 2005. Swimming speeds of larval coral reef fishes: Impacts on self-recruitment and dispersal. *Mar. Ecol. Prog. Ser.* 285, 223–232. <https://doi.org/10.3354/meps285223>
- Fithian, W., Hastie, T., 2013. Finite-sample equivalence in statistical models for presence-only data. *Ann. Appl. Stat.* 7, 1917–1939. <https://doi.org/10.1214/13-AOAS667>
- Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J.C., Smith-Vaniz, W.F., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcón, J.M., Bowen, B.W., Bernardi, G., 2008. Atlantic reef fish biogeography and evolution. *J. Biogeogr.* 35, 22–47. <https://doi.org/10.1111/j.1365-2699.2007.01790.x>
- Fodrie, F.J., Heck, K.L., Powers, S.P., Graham, W., Robinson, K., 2010. Climate-related, decadal-scale assemblage changes of seagrass-associated fishes in the northern Gulf of Mexico. *Glob. Chang. Biol.* 16, 48–59. <https://doi.org/10.1111/j.1365-2486.2009.01889.x>
- Fordham, D.A., Akçakaya, H.R., Brook, B.W., Rodríguez, A., Alves, P.C., Civantos, E., Triviño, M., Watts, M.J., Araújo, M.B., 2013. Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nat. Clim. Chang.* 3, 899–903. <https://doi.org/10.1038/nclimate1954>
- Fortin, M.J., Keitt, T.H., Maurer, B.A., Taper, M.L., Kaufman, D.M., Blackburn, T.M., 2005. Species' geographic ranges and distributional limits: Pattern analysis and statistical issues. *Oikos* 108, 7–17. <https://doi.org/10.1111/j.0030-1299.2005.13146.x>
- Fourcade, Y., Besnard, A.G., Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* 27, 245–256. <https://doi.org/10.1111/geb.12684>
- Fournier, A., Barbet-Massin, M., Rome, Q., and Courchamp, F. 2017. Predicting species distribution combining multi-scale drivers. *Glob. Ecol. Conserv.* 12, 215–226.
- Fournier, A., Barbet-Massin, M., Rome, Q., Courchamp, F. 2017. Predicting species distribution combining multi-scale drivers. *Global Ecology and Conservation*, 12: 215-226.
- Frias-Torres S. 2006. Habitat use of juvenile goliath grouper *Epinephelus itajara* in the Florida Keys, USA. *Endanger Species Res* 2:1–6.
- Frias-Torres, S., 2006. Habitat use of juvenile goliath grouper *Epinephelus itajara* in the Florida Keys, USA. *Endanger. Species Res.* 2, 1–6. <https://doi.org/10.3354/esr002001>
- Funk, V.A., Richardson, K.S., 2002. Systematic data in biodiversity studies: Use it or lose it. *Syst. Biol.* 51, 303–316. <https://doi.org/10.1080/10635150252899789>
- Funk, V.A., Zermoglio, M.F., Nasir, N., 1999. Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. *Biodivers. Conserv.* 8, 727–751. <https://doi.org/10.1023/A:1008877222842>

- Gagliano, M., McCormick, M.I., Meekan, M.G., 2007. Temperature-induced shifts in selective pressure at a critical developmental transition. *Oecologia* 152, 219–225. <https://doi.org/10.1007/s00442-006-0647-1>
- Gaston, K.J., 1991. How Large Is a Species' Geographic Range? *Oikos* 61, 434. <https://doi.org/10.2307/3545251>
- Gaston, K.J., Fuller, R.A., 2009. The sizes of species' geographic ranges. *J. Appl. Ecol.* 46, 1–9. <https://doi.org/10.1111/j.1365-2664.2008.01596.x>
- Gerhardinger, L.C., Hostim-Silva, M., Medeiros, R.P., Matarezi, J., Bertoncini, Á.A., Freitas, M.O., Ferreira, B.P., 2009. Fishers' resource mapping and goliath grouper *Epinephelus itajara* (Serranidae) conservation in Brazil. *Neotrop. Ichthyol.* 7, 93–102. <https://doi.org/10.1590/S1679-62252009000100012>
- Gibbons, D.W., Donald, P.F., Bauer, H.G., Fornasari, L., Dawson, I.K., 2007. Mapping avian distributions: The evolution of bird atlases. *Bird Study* 54, 324–334. <https://doi.org/10.1080/00063650709461492>
- Giglio, V.J., Adelir-Alves, J., Gerhardinger, L.C., Grecco, F.C., Daros, F.A., Bertoncini, Á.A., 2014a. Habitat use and abundance of goliath grouper *Epinephelus itajara* in Brazil: A participative survey. *Neotrop. Ichthyol.* 12, 803–810. <https://doi.org/10.1590/1982-0224-20130166>
- Giglio, V.J., Bertoncini, Á.A., Ferreira, B.P., Hostim-Silva, M., Freitas, M.O., 2014b. Landings of goliath grouper, *Epinephelus itajara*, in Brazil: Despite prohibited over ten years, fishing continues. *Nat. e Conserv.* 12, 118–123. <https://doi.org/10.1016/j.ncon.2014.09.004>
- Gilmore RG, Bullock LH, Berry FH (1978) Hypothermal mortality in marine fishes of south-central Florida. *Northeast Gulf Sci* 2:77–97
- Gilmore, R.G., Bullock, L.H., Berry, F.H., 1978. Hypothermal Mortality in Marine Fishes of South-Central Florida, January, 1977. *Northeast Gulf Sci.* 2. <https://doi.org/10.18785/negs.0202.01>
- Giri, C., Ochieng, E., Tieszen, L.L., Zhu, Z., Singh, A., Loveland, T., Masek, J., Duke, N., 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Glob. Ecol. Biogeogr.* 20, 154–159. <https://doi.org/10.1111/j.1466-8238.2010.00584.x>
- Gleckler, P. J. et al. 2012. Human-induced global ocean warming on multidecadal timescales. *Nature Climate Change*, 2(7), 524-529.
- Gonzalez, J., 2010. Nálisis de datos de 4–52. <https://doi.org/10.1111/j.2008.0030-1299.16434.x>
- Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA, et al. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*. 2008; 45(1):239-47.

- Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19, 497–503. <https://doi.org/10.1016/j.tree.2004.07.006>
- Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. *Glob. Ecol. Biogeogr.* 15, 578–587. <https://doi.org/10.1111/j.1466-8238.2006.00257.x>
- Graham, C.H., Moritz, C., Williams, S.E., 2006. Habitat history improves prediction of biodiversity in rainforest fauna. *Proc. Natl. Acad. Sci. U. S. A.* 103, 632–636. <https://doi.org/10.1073/pnas.0505754103>
- Graham, R.T., Rhodes, K.L., Castellanos, D., 2009. Characterization of the goliath grouper *Epinephelus itajara* fishery of southern Belize for conservation planning. *Endanger. Species Res.* 7, 195–204. <https://doi.org/10.3354/esr00187>
- Green, E.P., Short, F.T., 2004. World atlas of seagrasses, UNEP World Conservation Monitoring Centre. <https://doi.org/10.5860/choice.41-3160>
- Gueta, T. & Carmel, Y. 2016. Quantifying the value of user-level data cleaning for big data: A case study using mammal distribution models. *Ecological Informatics* 34: 139–145.
- Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 2005; 8: 993-1009.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., Mcdonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hagar, J.C., Yost, A., Haggerty, P.K., 2020. Incorporating LiDAR metrics into a structure-based habitat model for a canopy-dwelling species. *Remote Sens. Environ.* 236, 111499. <https://doi.org/10.1016/j.rse.2019.111499>
- Hageer, Y., Esperón-Rodríguez, M., Baumgartner, J. B, Beaumont, L. J. Climate, soil or both? Which variables are better predictors of the distributions of Australian shrub species? *PeerJ*, 5: e3446.
- Hales, S., Weinstein, P., Woodward, A., 1999. Ciguatera (fish poisoning), El Niño, and Pacific Sea surface temperatures. *Ecosyst. Heal.* 5, 20–25. <https://doi.org/10.1046/j.1526-0992.1999.09903.x>
- Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., Lowndes, J.S., Rockwood, R.C., Selig, E.R., Selkoe, K.A., Walbridge, S., 2015. Spatial and

temporal changes in cumulative human impacts on the world's ocean. *Nat. Commun.* 6, 1–7. <https://doi.org/10.1038/ncomms8615>

Halpern, B.S., Gaines, S.D., Warner, R.R., 2005. Habitat size, recruitment, and longevity as factors limiting population size in stage-structured species. *Am. Nat.* 165, 82–94. <https://doi.org/10.1086/426672>

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., Watson, R., 2008. A global map of human impact on marine ecosystems. *Science* (80-.). 319, 948–952. <https://doi.org/10.1126/science.1149345>

Hannemann, H., Willis, K. J., Macias-Fauria, M. 2016. The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. *Global Ecology and Biogeography*, 25, 26–35.

Harrison, H.B., Williamson, D.H., Evans, R.D., Almany, G.R., Thorrold, S.R., Russ, G.R., Feldheim, K.A., Van Herwerden, L., Planes, S., Srinivasan, M., Berumen, M.L., Jones, G.P., 2012. Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Curr. Biol.* 22, 1023–1028. <https://doi.org/10.1016/j.cub.2012.04.008>

Harvell, C.D., Mitchell, C.E., Ward, J.R., Altizer, S., Dobson, A.P., Ostfeld, R.S., Samuel, M.D., 2002. Climate warming and disease risks for terrestrial and marine biota. *Science* (80-.). 296, 2158–2162. <https://doi.org/10.1126/science.1063699>

Hazel, J.R., Prosser, C.L., 1974. Molecular mechanisms of temperature compensation in poikilotherms. *Physiol. Rev.* 54, 620–677. <https://doi.org/10.1152/physrev.1974.54.3.620>

Hazen EL, Jorgensen S, Rykaczewski RR, Bograd SJ, Foley DG, Jonsen ID, et al. Predicted habitat shifts of Pacific top predators in a changing climate. *Nat Clim Chang.* 2013; 3:234±238.

Heck, K.L., Hays, G., Orth, R.J., 2003. Critical evaluation of the nursery role hypothesis for seagrass meadows. *Mar. Ecol. Prog. Ser.* 253, 123–136. <https://doi.org/10.3354/meps253123>

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, 30: 751-777.

Hernandez PA, Graham CH, Master LL, Albert DL. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography (Cop)*. 2006; 29:773-785.

Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 183–195. <https://doi.org/10.1098/rstb.2003.1388>

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

- Hijmans, R. J., & Hall, W. (2016). Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology*, 93, 679–688.
- Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. *dismo*: species distribution modeling. – R package ver. 1.1–4 (2016).
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2): 142–152.
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Modell.* 199, 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hofmann, G.E., Barry, J.P., Edmunds, P.J., Gates, R.D., Hutchins, D.A., Klinger, T., Sewell, M.A., 2010. The Effect of Ocean Acidification on Calcifying Organisms in Marine Ecosystems: An Organism-to-Ecosystem Perspective. *Annu. Rev. Ecol. Evol. Syst.* 41, 127–147. <https://doi.org/10.1146/annurev.ecolsys.110308.120227>
- Holt, R.D., 2003. On the evolutionary ecology of species' ranges. *Evol. Ecol. Res.* 5, 159–178.
- Houde, E.D., 1989. Comparative growth, mortality, and energetics of marine fish larvae: temperature and implied latitudinal effects. *Fish. Bull.* 87, 471–495.
- Hu, J., Hu, H., Jiang, Z., 2010. The impacts of climate change on the wintering distribution of an endangered migratory bird. *Oecologia* 164, 555–565. <https://doi.org/10.1007/s00442-010-1732-z>
- Hughes, T.P., Bellwood, D.R., Connolly, S.R., 2002. Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecol. Lett.* 5, 775–784. <https://doi.org/10.1046/j.1461-0248.2002.00383.x>
- Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13384–13389. <https://doi.org/10.1073/pnas.0704469104>
- IPBES. 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. Díaz S, Settele J, Brondizio ES, Ngo HT, Guèze M, Agard J, Arneth A, Balvanera P, Brauman KA, Butchart SHM, et al., editors. Bonn (Germany): IPBES secretariat.
- IPCC (2014) Summary for Policymakers. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland Cambridge, 1–32. <https://doi.org/10.1017/CBO9781107415324>
- IPCC (2018) Summary for Policymakers. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts*

to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.

Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., Gutiérrez, J. M. Gutiérrez. 2015. A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling*, 312:166–174.

IUCN. 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Prepared by the Standards and Petitions Committee. Downloadable from <http://www.iucnredlist.org/documents/RedListGuidelines.pdf>.

Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B., Lenihan, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J., Warner, R.R., 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science* (80-). 293, 629–637. <https://doi.org/10.1126/science.1059199>

Janžekovič F, Novak T. 2012. PCA – A Powerful Method for Analyze Ecological Niches. In: Sanguansat P, editor editors. *Principal Component Analysis - multidisciplinary Applications*. 127–142.

Jarnevich, C. S., Stohlgren, T. J., Kumar, S., Morissette, J. T., & Holcombe, T. R. 2015. Caveats for correlative species distribution modeling. *Ecological Informatics*, 29: 6–15.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology and Evolution*, 27: 151–159.

Jetz, W., Sekercioglu, C.H., Watson, J.E.M., 2008. Ecological correlates and conservation implications of overestimating species geographic ranges. *Conserv. Biol.* 22, 110–119. <https://doi.org/10.1111/j.1523-1739.2007.00847.x>

Jin, J. & Yang, J. 2020. BDCleaner: a workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. *Global Ecology and Conservation*, 21: e00852.

Jueterbock, A., Smolina, I., Coyer, J.A., Hoarau, G., 2016. The fate of the Arctic seaweed *Fucus distichus* under climate change: An ecological niche modeling approach. *Ecol. Evol.* 6, 1712–1724. <https://doi.org/10.1002/ece3.2001>

Jueterbock, A., Tyberghein, L., Verbruggen, H., Coyer, J.A., Olsen, J.L., Hoarau, G., 2013. Climate change impact on seaweed meadow distribution in the North Atlantic rocky intertidal. *Ecol. Evol.* 3, 1356–1373. <https://doi.org/10.1002/ece3.541>

Kadmon, R. et al. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* 14: 401–413.

Kadmon, R., Farber, O., Danin, A., 2009. Effect of Roadside Bias on the Accuracy of Predictive Maps Produced by Bioclimatic Models Published by : Ecological Society of America America (NY). 14, 401–413.

- 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. <https://doi.org/10.3354/meps316285>
- Keller, C.F., 2007. Global warming 2007 an update to global warming: The balance of evidence and its policy implications. *ScientificWorldJournal.* 7, 381–399. <https://doi.org/10.1100/tsw.2007.91>
- Kennedy, V.S., 1990. Anticipated Effects of Climate Change on Estuarine and Coastal Fisheries. *Fisheries* 15, 16–24. [https://doi.org/10.1577/1548-8446\(1990\)015<0016:aeocco>2.0.co;2](https://doi.org/10.1577/1548-8446(1990)015<0016:aeocco>2.0.co;2)
- Kimura, R., Takami, H., Ono, T., Onitsuka, T., Nojiri, Y., 2011. Effects of elevated pCO₂ on the early development of the commercially important gastropod, Ezo abalone *Haliotis discus hannai*. *Fish. Oceanogr.* 20, 357–366. <https://doi.org/10.1111/j.1365-2419.2011.00589.x>
- Kleisner KM, Fogarty MJ, McGee S, Hare JA, Moret S, Perretti CT, et al. Marine species distribution shifts on the U.S. northeast continental shelf under continued ocean warming. *Prog Oceanog.* 2017; 153:24±36.
- Knowlton, N., 2001. Coral Reef Biodiversity-Habitat Size Matters 292.
- Koenig, C.C., Bueno, L.S., Coleman, F.C., Cusick, J.A., Ellis, R.D., Kingon, K., Locascio, J.V., Malinowski, C., Murie, D.J. and Stallings, C.D. (2016). Diel, lunar, and seasonal spawning patterns of the Atlantic goliath grouper, *Epinephelus itajara*, off Florida, United States. *Bulletin of Marine Science* 93(2): 391-406.
- Koenig, C.C., Coleman, F.C., Eklund, A.M., Schull, J., Ueland, J., 2007. Mangroves as essential nursery hábitat for goliath grouper (*Epinephelus itajara*). *Bull. Mar. Sci.* 80, 567–586.
- Koenig, C.C., Coleman, F.C., Kingon, K. 2011. Pattern of recovery of the Goliath Grouper *Epinephelus itajara* population in the Southeastern US. *Bulletin of Marine Science* 87(4): 891-911.
- Koenig, C.C., Coleman, F.C., Kingon, K.C., 2009. Recovery of the Goliath Grouper (*Epinephelus itajara*) Population of the Southeastern U . S . Recuperación de la Población del Mero Goliath (*Epinephelus itajara*) del Sureste de los Estados Unidos Déplacement et Modeles de Distribution de la Population d.
- Kroeker, K.J., Kordas, R.L., Crim, R., Hendriks, I.E., Ramajo, L., Singh, G.S., Duarte, C.M., Gattuso, J.P., 2013. Impacts of ocean acidification on marine organisms: Quantifying sensitivities and interaction with warming. *Glob. Chang. Biol.* 19, 1884–1896. <https://doi.org/10.1111/gcb.12179>
- Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13, 66–69.
- Lara, M.R., Schull, J., Jones, D.L., Allman, R., 2009. Early life history stages of goliath grouper *Epinephelus itajara* (Pisces: Epinephelidae) from Ten Thousand Islands, Florida. *Endanger. Species Res.* 7, 221–228. <https://doi.org/10.3354/esr00193>

- Last, P.R., White, W.T., Gledhill, D.C., Hobday, A.J., Brown, R., Edgar, G.J., Pecl, G., 2011. Long-term shifts in abundance and distribution of a temperate fish fauna: A response to climate change and fishing practices. *Glob. Ecol. Biogeogr.* 20, 58–72. <https://doi.org/10.1111/j.1466-8238.2010.00575.x>
- Lawler, J.J., White, D., Neilson, R.P., Blaustein, A.R., 2006. Predicting climate-induced range shifts: Model differences and model reliability. *Glob. Chang. Biol.* 12, 1568–1584. <https://doi.org/10.1111/j.1365-2486.2006.01191.x>
- Lawton, J.H., 1999. Are There General Laws in Ecology? *Oikos* 84, 177. <https://doi.org/10.2307/3546712>
- Leis, J.M., Siebeck, U., Dixon, D.L., 2011. How nemo finds home: The neuroecology of dispersal and of population connectivity in larvae of marine fishes. *Integr. Comp. Biol.* 51, 826–843. <https://doi.org/10.1093/icb/icr004>
- Leitão, P. J., Santos, M. J. 2019. Improving models of species ecological niches: a remote sensing overview. *Frontiers in Ecology and Evolution*, 7: 9.
- Lejeusne, C., Chevaldonné, P., Pergent-Martini, C., Boudouresque, C.F., Pérez, T., 2010. Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends Ecol. Evol.* 25, 250–260. <https://doi.org/10.1016/j.tree.2009.10.009>
- Leroy, B. et al. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography* 45(9): 1994–2002.
- Leroy, B., Delsol, R., Hugué, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M., Bellard, C., 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *J. Biogeogr.* 45, 1994–2002. <https://doi.org/10.1111/jbi.13402>
- Lewis, S. L., Maslin, M. A. (2015) Defining the Anthropocene. *Nature* 519, 171–180.
- Lin, Y. et al. 2015. Uncertainty analysis of crowd-sourced and professionally collected field data used in species distribution models of Taiwanese moths. - *Biological Conservation* 248 181: 102–110.
- Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* 40, 778–789. <https://doi.org/10.1111/jbi.12058>
- Liu, W., Xie, S.P., Liu, Z., Zhu, J., 2017. Overlooked possibility of a collapsed atlantic meridional overturning circulation in warming climate. *Sci. Adv.* 3, 1–8. <https://doi.org/10.1126/sciadv.1601666>
- Lloyd, P., Plagányi, É.E., Weeks, S.J., Magno-Canto, M., Plagányi, G., 2012. Ocean warming alters species abundance patterns and increases species diversity in an African sub-tropical reef-fish community. *Fish. Oceanogr.* 21, 78–94. <https://doi.org/10.1111/j.1365-2419.2011.00610.x>

- Loaiciga, H.A., Valdes, J.B., Vogel, R., Garvey, J., Schwarz, H., 1996. Global warming and the hydrologic cycle. *J. Hydrol.* 174, 83–127. [https://doi.org/10.1016/0022-1694\(95\)02753-X](https://doi.org/10.1016/0022-1694(95)02753-X)
- Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B., Ackerly, D.D., 2009. The velocity of climate change. *Nature* 462, 1052–1055. <https://doi.org/10.1038/nature08649>
- Lobato, C.M.C., Soares, B.E., Begot, T.O.R., Montag, L.F. de A., 2016. Tidal pools as habitat for juveniles of the goliath grouper *Epinephelus itajara* (Lichtenstein 1822) in the Amazonian coastal zone, Brazil. *Nat. e Conserv.* 14, 20–23. <https://doi.org/10.1016/j.ncon.2015.12.001>
- Loiselle, B. A. et al. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35: 105–116.
- Louppe, V., Leroy, B., Herrel, A., Veron, G., 2019. Current and future climatic regions favourable for a globally introduced wild carnivore, the raccoon *Procyon lotor*. *Sci. Rep.* 9, 1–13. <https://doi.org/10.1038/s41598-019-45713-y>
- Loyola, R.D., Lemes, P., Faleiro, F. V., Trindade-Filho, J., Machado, R.B., 2012. Severe Loss of Suitable Climatic Conditions for Marsupial Species in Brazil: Challenges and Opportunities for Conservation. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0046257>
- Lozier, J.D.; Aniello, P.; Hickerson, M.J. Predicting the distribution of Sasquatch in western North America: Anything goes with ecological niche modelling. *J. Biogeogr.* 2009, 36,1623–1627.
- Macintosh, D.J., Ashton, E.C., 2002. A Review of Mangrove Biodiversity Conservation and Management. *Ecosystems*.
- Madin, E.M.P., Ban, N.C., Doubleday, Z.A., Holmes, T.H., Pecl, G.T., Smith, F., 2012. Socio-economic and management implications of range-shifting species in marine systems. *Glob. Environ. Chang.* 22, 137–146. <https://doi.org/10.1016/j.gloenvcha.2011.10.008>
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases. *Global Ecology and Biogeography*, 24(8), 973–984.
- Malinowski, C.R., 2019. High mercury concentrations in Atlantic Goliath Grouper: Spatial analysis of a vulnerable species. *Mar. Pollut. Bull.* 143, 81–91. <https://doi.org/10.1016/j.marpolbul.2019.04.006>
- Maragos, J.E., Crosby, M.P., McManus, J.W., 1996. Coral reefs and biodiversity: A critical and threatened relationship. *Oceanography* 9, 83–99. <https://doi.org/10.5670/oceanog.1996.31>

- Marcelino, V.R., Verbruggen, H., 2015. Ecological niche models of invasive seaweeds. *J. Phycol.* 51, 606–620. <https://doi.org/10.1111/jpy.12322>
- Marini MAÃ , Barbet-Massin M, Lopes LE, Jiguet F. Predicting the occurrence of rare Brazilian birds with species distribution models. *J Ornithol.* 2010; 151: 857-866.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15, 59–69.
- Mateo RG, FelicÃsimo M, MuÃoz J. Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. *J Veg Sci.* 2010; 21: 908-922.
- McKelvey, K.S.; Aubry, K.B.; Schwartz, M.K. Using anecdotal occurrence data for rare or elusive species: The illusion of reality and a call for evidentiary standards. *Bioscience* 2008, 58, 549–555.
- McLeod, I.M., Jones, R.E., Jones, G.P., Takahashi, M., McCormick, M.I., 2015. Interannual variation in the larval development of a coral reef fish in response to temperature and associated environmental factors. *Mar. Biol.* 162, 2379–2389. <https://doi.org/10.1007/s00227-015-2765-y>
- MEA - Millennium Ecosystem Assessment . 2005. *Ecosystems and Human Well Being: Synthesis*. Island Press, Washington DC.
- Melo-Merino, S.M., Reyes-Bonilla, H., Lira-Noriega, A., 2020. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecol. Modell.* 415, 108837. <https://doi.org/10.1016/j.ecolmodel.2019.108837>
- Meyer, C., 2016. Limitations in global information on species occurrences. *Front. Biogeogr.* 8. <https://doi.org/10.21425/f5fbg28195>
- Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* 6, 1–8. <https://doi.org/10.1038/ncomms9221>
- Meyer, C., Weigelt, P., Kreft, H., 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006. <https://doi.org/10.1111/ele.12624>
- Miller, G.M., Watson, S.A., Donelson, J.M., McCormick, M.I., Munday, P.L., 2012. Parental environment mediates impacts of increased carbon dioxide on a coral reef fish. *Nat. Clim. Chang.* 2, 858–861. <https://doi.org/10.1038/nclimate1599>
- Milly, P.C.D., Wetherald, R.T., Dunne, K.A., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. *Nature* 415, 514–517. <https://doi.org/10.1038/415514a>
- Minsky, E.M., Lorini, M.L., Arlé, C.E., 2017. Dinâmica da distribuição e estabilidade das áreas ambientalmente adequadas para o mero, *Epinephelus itajara* (Lichtenstein,

1822) (Perciformes, Epinephelidae), em cenários pretéritos de mudanças climáticas do quaternário até o presente.

Mireille, C., Gatti, C. M., Roué, M., Darius, T. 2020. Chapter 11. Ciguatera-Causing Dinoflagellates in the Genera *Gambierdisus* and *Fukuyoa*: Distribution, Ecophysiology and Toxicology. In Rao, D.V.S. (ed.) *Dinoflagellates: Classification, Evolution, Physiology and Ecological Significance*. Nova Science Publishers Inc. Pp.405-457.

Moerman, D.E., Estabrook, G.F., 2006. The botanist effect: Counties with maximal species richness tend to be home to universities and botanists. *J. Biogeogr.* 33, 1969–1974. <https://doi.org/10.1111/j.1365-2699.2006.01549.x>

Molinos JG, Halpern BS, Schoeman DS, Brown CJ, Kiessling W, et al. 2015. Climate velocity and the future global redistribution of marine biodiversity. *Nat. Clim. Change* 6:83–88.

Monzón, J., Moyer-Horner, L., Palamar, M.B., 2011. Climate Change and Species Range Dynamics in Protected Areas. *Bioscience* 61, 752–761. <https://doi.org/10.1525/bio.2011.61.10.5>

Morley JW, Selden RL, Latour RJ, Frölicher TL, Seagraves RJ, Pinsky ML. 2018. Projecting shifts in thermal habitat for 686 species on the North American continental shelf. *PLOS ONE* 13:e0196127.

Moss, R. H. et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463: 747–756.

Mullon, C., Cury, P., Penven, P., 2002. Evolutionary individual-based model for the recruitment of anchovy (*Engraulis capensis*) in the southern Benguela. *Can. J. Fish. Aquat. Sci.* 59, 910–922. <https://doi.org/10.1139/f02-064>

Munday, P.L., Donelson, J.M., Dixson, D.L., Endo, G.G.K., 2009. Effects of ocean acidification on the early life history of a tropical marine fish. *Proc. R. Soc. B Biol. Sci.* 276, 3275–3283. <https://doi.org/10.1098/rspb.2009.0784>

Munday, P.L., Gagliano, M., Donelson, J.M., Dixson, D.L., Thorrold, S.R., 2011. Ocean acidification does not affect the early life history development of a tropical marine fish. *Mar. Ecol. Prog. Ser.* 423, 211–221. <https://doi.org/10.3354/meps08990>

Munday, P.L., Jones, G.P., 1998. Ecological implications of small body size among coral-reef fishes. *Oceanogr. Mar. Biol.*

Murray, M.R., Zisman, S.A., Furley, P.A., Munro, D.M., Gibson, J., Ratter, J., Bridgewater, S., Minty, C.D., Place, C.J., 2003. The mangroves of Belize Part 1. distribution, composition and classification. *For. Ecol. Manage.* 174, 265–279. [https://doi.org/10.1016/S0378-1127\(02\)00036-1](https://doi.org/10.1016/S0378-1127(02)00036-1)

Naimi, B., Araújo, M.B., 2016. Sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography (Cop.)*. 39, 368–375. <https://doi.org/10.1111/ecog.01881>

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog. Phys. Geogr.* 34, 3–22. <https://doi.org/10.1177/0309133309355630>

Nishimura, T., Sato, S., Tawong, W., Sakanari, H., Uehara, K., Shah, M.M.R., Suda, S., Yasumoto, T., Taira, Y., Yamaguchi, H., Adachi, M., 2013. Genetic Diversity and Distribution of the Ciguatera-Causing Dinoflagellate *Gambierdiscus* spp. (Dinophyceae) in Coastal Areas of Japan. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0060882>

Oreskes, N., 2005. The Scientific Consensus on Climate Change. *Science* (80-.). 306, 2004–2005. <https://doi.org/10.1126/science.1103618>

Orth, R.J., Carruthers, T.J.B., Dennison, W.C., Duarte, C.M., Fourqurean, J.W., Heck, K.L., Hughes, A.R., Kendrick, G.A., Kenworthy, W.J., Olyarnik, S., Short, F.T., Waycott, M., Williams, S.L., 2006. A Global Crisis for Seagrass Ecosystems. *Bioscience* 56, 987. [https://doi.org/10.1641/0006-3568\(2006\)56\[987:agcfse\]2.0.co;2](https://doi.org/10.1641/0006-3568(2006)56[987:agcfse]2.0.co;2)

Osborne PE, Leitão PJ. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Diversity and Distributions*. 2009; 15(4):671-81.

Pace, D.S., Arcangeli, A., Mussi, B., Vivaldi, C., Ledon, C., Lagorio, S., Giacomini, G., Pavan, G., Ardizzone, G., 2018. Habitat suitability modeling in different sperm whale social groups. *J. Wildl. Manage.* 82, 1062–1073. <https://doi.org/10.1002/jwmg.21453>

Palmer, M. D., Good, S. A., Haines, K., Rayner, N. A., Stott, P. A. 2009. A new perspective on warming of the global oceans. *Geophysical Research Letters*, 36: L20709.

Papeş M, Gaubert P. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Divers Distrib.* 2007; 13: 890-902.

Parker, L.M., Ross, P.M., O'Connor, W.A., 2009. The effect of ocean acidification and temperature on the fertilization and embryonic development of the Sydney rock oyster *Saccostrea glomerata* (Gould 1850). *Glob. Chang. Biol.* 15, 2123–2136. <https://doi.org/10.1111/j.1365-2486.2009.01895.x>

Parmesan, C., 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annu. Rev. Ecol. Evol. Syst.* 37, 637–669. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110100>

Pearson, R. G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34: 102–117.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34, 102–117. <https://doi.org/10.1111/j.1365-2699.2006.01594.x>

- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. *J. Biogeogr.* 33, 1704–1711. <https://doi.org/10.1111/j.1365-2699.2006.01460.x>
- Pena, J.C. de C., Kamino, L.H.Y., Rodrigues, M., Mariano-Neto, E., de Siqueira, M.F., 2014. Assessing the conservation status of species with limited available data and disjunct distribution. *Biol. Conserv.* 170, 130–136. <https://doi.org/10.1016/j.biocon.2013.12.015>
- Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A. 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from crosscontinental plant invasions. *Global Ecol Biogeogr.* 26(3), 275–287.
- Peterson AT, Cobos ME, Jiménez-García D. (2018). Major challenges for correlational ecological niche model projections to future climate conditions. *Annals of the New York Academy of Sciences*, 1429(1):66-77.
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB (2011) *Ecological niches and geographic distributions*. Princeton University Press, Princeton.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E, Nakamura, M., A., Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Bastos Araujo, M., 2011. *Ecological niches and geographic distributions*, Princeton University Press. <https://doi.org/10.5860/CHOICE.49-6266>
- Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: An example with *Solenopsis invicta* and *Solenopsis richteri*. *Glob. Ecol. Biogeogr.* 17, 135–144. <https://doi.org/10.1111/j.1466-8238.2007.00347.x>
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. *Ecological Niches and Geographic Distributions*, *Ecological Niches and Geographic Distributions (MPB-49)*. <https://doi.org/10.23943/princeton/9780691136868.001.0001>
- Phillips, B.F., Pérez-Ramírez, M., 2018. *Climate Change Impacts on Fisheries and Aquaculture*.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. <https://doi.org/10.1890/07-2153.1>
- Pierce, D. W., Gleckler, P. J., Barnett, T. P., Santer, B. D., & Durack, P. J. 2012. The fingerprint of human induced changes in the ocean's salinity and temperature fields. *Geophysical Research Letters*, 39:: L21704.
- Pinsky ML, Eikeset AM, McCauley DJ, Payne JL, Sunday JM. 2019. Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* 569:108–111.

Pinsky, M. L., Selden, R. L., Zoë, J. 2020. KitchelClimate-Driven Shifts in Marine Species Ranges: Scaling from Organisms to Communities. *Annual Review of Marine Science*, 12:153–79.

Poloczanska, E.S., Brown, C.J., Sydeman, W.J., Kiessling, W., Schoeman, D.S., Moore, P.J., Brander, K., Bruno, J.F., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Holding, J., Kappel, C. V., O’Connor, M.I., Pandolfi, J.M., Parmesan, C., Schwing, F., Thompson, S.A., Richardson, A.J., 2013. Global imprint of climate change on marine life. *Nat. Clim. Chang.* 3, 919–925. <https://doi.org/10.1038/nclimate1958>

Pörtner, H.O., 2002. Climate variations and the physiological basis of temperature dependent biogeography: Systemic to molecular hierarchy of thermal tolerance in animals. *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* 132, 739–761. [https://doi.org/10.1016/S1095-6433\(02\)00045-4](https://doi.org/10.1016/S1095-6433(02)00045-4)

Pörtner, H.O., Farrell, A.P., 2008. Ecology: Physiology and climate change. *Science* (80-.). 322, 690–692. <https://doi.org/10.1126/science.1163156>

Pörtner, H.O., Knust, R., 2007. Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science* (80-.). 315, 95–97. <https://doi.org/10.1126/science.1135471>

Qiao, H., Soberón, J., & Peterson, A. T. (2015). No silver bullets in correlative ecological niche modelling: Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, 6(10), 1126–1136.

Radosavljevic, A., Anderson, R. P. 2014. Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, 41, 629–643.

Rahmstorf, S., 2013. Thermohaline Circulation. *Encycl. Quat. Sci. Second Ed.* 737–747. <https://doi.org/10.1016/B978-0-444-53643-3.00020-0>

Reddy, S., Dávalos, L.M., 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *J. Biogeogr.* 30, 1719–1727. <https://doi.org/10.1046/j.1365-2699.2003.00946.x>

Regos, A., Gagne, L., Alcaraz-Segura, D. et al. (2019). Effects of species traits and environmental predictors on performance and transferability of ecological niche models. *Sci Rep* 9, 4221.

Reichert, B. K., Schnur, R., Bengtsson, L. 2002. Global ocean warming tied to anthropogenic forcing. *Geophysical Research Letters*, 29: 20–4.

Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics* 69, 274–281. <https://doi.org/10.1111/j.1541-0420.2012.01824.x>

Reside, A. E., Watsonb, I., VanDerWal, J., Kuttb, A. S. 2011. Incorporating low-resolution historic species location data decreases performance of distribution models. *Ecological Modelling*, 222: 3444– 3448.

- Reside, A.E., Watson, I., VanDerWal, J., Kutt, A.S., 2011. Incorporating low-resolution historic species location data decreases performance of distribution models. *Ecol. Modell.* 222, 3444–3448. <https://doi.org/10.1016/j.ecolmodel.2011.06.015>
- Rhein, M. et al. 2013. Observations: ocean. In Stocker, T. F. et al. (eds) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press, pp. 255–316.
- Ritz, D.A., 1994. Social Aggregation in Pelagic Invertebrates. *Adv. Mar. Biol.* 30, 155–216. [https://doi.org/10.1016/S0065-2881\(08\)60063-2](https://doi.org/10.1016/S0065-2881(08)60063-2)
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Warton, D. I. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Robertson, A.I., 1986. Leaf-burying crabs: Their influence on energy flow and export from mixed mangrove forests (*Rhizophora* spp.) in northeastern Australia. *J. Exp. Mar. Bio. Ecol.* 102, 237–248. [https://doi.org/10.1016/0022-0981\(86\)90179-6](https://doi.org/10.1016/0022-0981(86)90179-6)
- Robertson, A.I., Blaber, S.J.M., 1992. Plankton, epibenthos and fish communities 41, 173–224. <https://doi.org/10.1029/ce041p0173>
- Robinson LM, Hobday AJ, Possingham HP, Richardson AJ. Trailing edges projected to move faster than leading edges for large pelagic fish habitats under climate change. *Deep-sea Res II.* 2015; 113:225±234.
- 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. <https://doi.org/10.1111/j.1466-8238.2010.00636.x>
- Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E., Lundquist, C.J., 2017. A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. *Front. Mar. Sci.* 4, 1–11. <https://doi.org/10.3389/fmars.2017.00421>
- Roessig, J.M., Woodley, C.M., Cech Jr, J.J., Hansen, L.J., 2004. Effects of global climate change on marine and estuarine fishes and fisheries. *Rev. Fish Biol. Fish.* 14, 251–275. <https://doi.org/10.1016/j.jmaa.2014.03.081>
- Rombouts, I., Beaugrand, G., Dauvin, J.C., 2012. Potential changes in benthic macrofaunal distributions from the English Channel simulated under climate change scenarios. *Estuar. Coast. Shelf Sci.* 99, 153–161. <https://doi.org/10.1016/j.ecss.2011.12.026>
- Rondinini, C., Boitani, L., 2007. Systematic conservation planning and the cost of tackling conservation conflicts with large carnivores in Italy. *Conserv. Biol.* 21, 1455–1462. <https://doi.org/10.1111/j.1523-1739.2007.00834.x>
- Rondinini, C., Stuart, S., Boitani, L., 2005. Habitat suitability models and the shortfall in conservation planning for African vertebrates. *Conserv. Biol.* 19, 1488–1497. <https://doi.org/10.1111/j.1523-1739.2005.00204.x>

- 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. <https://doi.org/10.1111/j.1461-0248.2006.00970.x>
- Roubicek AJ, VanDerWal J, Beaumont LJ, Pitman AJ, Wilson P, Hughes L. 2010. Does the choice of climate baseline matter in ecological niche modelling? *Ecological Modelling* 221: 2280–2286.
- Roughgarden, J., 2009. Is there a general theory of community ecology? *Biol. Philos.* 24, 521–529. <https://doi.org/10.1007/s10539-009-9164-z>
- Russo, D., Maiorano, L., Rebelo, H., Preatoni, D.G., 2016. Species distribution models: Exploring patterns and processes in mammal ecology. *Hystrix* 27, 3–4. <https://doi.org/10.4404/hystrix-27.1-12101>
- S., B.M., R., V.P., E., N.S., A., S.F.K., 2002. Evaluating resource selection functions. *Ecol. Modell.* 157, 281–300.
- Sadovy, Y., Eklund, a. M., 1999. Synopsis of biological data on the Nassau grouper, *Epinephelus striatus* (Bloch, 1792), and the jewfish, *E. itajara* (Lichenstein, 1822). *Fao Fish. Synopsis* 68.
- Sandman, A.N., Wikström, S. A., Blomqvist, M., Kautsky, H., & Isaeus, M. (2013). Scale-dependent influence of environmental variables on species distribution: A case study on five coastal benthic species in the Baltic Sea. *Ecography*, 36, 354–363.
- Sasekumar, A., Chong, V.C., Leh, M.U., Cruz, R.D., Murray, J., Sala, E., Aburto-oropeza, O., Ezcurra, E., Danemann, G., Spalding, M., 1992. the Role of Mangroves in the Role of Mangroves in 195–201.
- Sbrocco, E.J., Barber, P.H., 2013. MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology* 94, 979. <https://doi.org/10.1890/12-1358.1>
- Seebacher, F., 2009. Responses to temperature variation: Integration of thermoregulation and metabolism in vertebrates. *J. Exp. Biol.* 212, 2885–2891. <https://doi.org/10.1242/jeb.024430>
- Sherman, R.E., Fahey, T.J., Howarth, R.W., 1998. Forest : in a Neotropical Interactions Mangrove and Sulfur Dynamics Iron , Phosphorus. *Oecologia* 115, 553–563. <https://doi.org/10.1007/s004420050553>
- Shideler, G.S., Sagarese, S.R., Harford, W.J. et al. Assessing the suitability of mangrove habitats for juvenile Atlantic goliath grouper. *Environ Biol Fish* 98, 2067–2082 (2015).
- Silva, D. P. et al. Seeking the flowers for the bees: Integrating biotic interactions into niche models to assess the distribution of the exotic bee species *Lithurgus huberi* in South America. *Ecol. Modell.* 273, 200–209 (2014).
- Silvy, Y., Guilyardi, E., Sallée, J., Durack, P. J. 2020. Human-induced changes to the global ocean water masses and their time of emergence, *Nature Climate Change*. <https://doi.org/10.1038/s41558-020-0878-x>

Slim, F.J., Hemminga, M.A., Ochieng, C., Jannink, N.T., Cocheret De La Morinière, E., Van Der Velde, G., 1997. Leaf litter removal by the snail *Terebralia palustris* (Linnaeus) and sesarmid crabs in an East African mangrove forest (Gazi Bay, Kenya). *J. Exp. Mar. Bio. Ecol.* 215, 35–48. [https://doi.org/10.1016/S0022-0981\(97\)00029-4](https://doi.org/10.1016/S0022-0981(97)00029-4)

Smith, T.J., 1987. Seed predation in relation to tree dominance and distribution in mangrove forests ' 68, 266–273.

Soberón, J.M. & Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359, 689– 698.

Soberón, J.M., Llorente, J.B., Oñate, L., 2000. The use of specimen-label databases for conservation purposes: An example using Mexican Papilionid and Pierid butterflies. *Biodivers. Conserv.* 9, 1441–1466. <https://doi.org/10.1023/A:1008987010383>

Sofaer, H.R., Jarnevich, C.S., Pearse, I.S., Smyth, R.L., Auer, S., Cook, G.L., Edwards, T.C., Guala, G.F., Howard, T.G., Morissette, J.T., Hamilton, H., 2019. Development and Delivery of Species Distribution Models to Inform Decision-Making. *Bioscience* 69, 544–557. <https://doi.org/10.1093/biosci/biz045>

Somodi, I., Lepesi, N., & Botta-Dukát, Z. 2017. Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecology and Evolution*, 7, 863–872.

Sorte, C.J.B., Williams, S.L., Carlton, J.T., 2010. Marine range shifts and species introductions: Comparative spread rates and community impacts. *Glob. Ecol. Biogeogr.* 19, 303–316. <https://doi.org/10.1111/j.1466-8238.2009.00519.x>

Soultan A, Safi K (2017) The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialisation. *PLoS ONE* 12(11): e0187906.

Sousa-Baena, M.S., Garcia, L.C., Peterson, A.T., 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Divers. Distrib.* 20, 369–381. <https://doi.org/10.1111/ddi.12136>

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *Bioscience* 57, 573. <https://doi.org/10.1641/B570707>

Spiers, A.G., 1999. Review of international/ continental wetland resources. *Glob. Rev. Wetl. Resour. priorities Wetl. Invent.* 63–104.

Stevenson, J.C., Kearney, M.S., Koch, E.W., 2002. Impacts of sea level rise on tidal wetlands and shallow water habitats: A case study from Chesapeake Bay. *Am. Fish. Soc. Symp.* 2002, 23–36.

Stockwell DR., Peterson AT. Effects of sample size on accuracy of species distribution models. *Ecol. Modell.* 2002; 148: 1-13.

- Sundby, S., Nakken, O., 2008. Spatial shifts in spawning habitats of Arcto-Norwegian cod related to multidecadal climate oscillations and climate change. *ICES J. Mar. Sci.* 65, 953–962. <https://doi.org/10.1093/icesjms/fsn085>
- Swets, J., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Synes, N. W., Osborne, P. E. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Glob. Ecol. Biogeogr.*, 20: 904–914.
- Syphard, A.D., Franklin, J. 2009. Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography*, 32: 907e918.
- Thorrold, S.R., Latkoczy, C., Swart, P.K., Jones, C.M., 2001. Natal Homing in a Marine Fish. *Science* (80-.). 291, 297–299. <https://doi.org/10.1126/science.291.5502.297>
- Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. *Glob. Chang. Biol.* 10, 2020–2027. <https://doi.org/10.1111/j.1365-2486.2004.00859.x>
- Thuiller, W., Guéguen, M., Renaud, J. et al. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nat Commun* 10, 1446.
- Thuiller, W., Richardson, D.M., Py Ek, P., Midgley, G.F., Hughes, G.O., Rouget, M., 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob. Chang. Biol.* 11, 2234–2250. <https://doi.org/10.1111/J.1365-2486.2005.01018.X>
- Tokarska, K. B., Hegerl, G. C., Schurer, A. P., Ribes, A., & Fasullo, J. T. 2019. Quantifying human contributions to past and future ocean warming and thermohaline sea level rise. *Environmental Research Letters*, 14:, 074020.
- Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* 7:9132.
- Tulloch, A.I.T. et al. 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.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O., 2012. Bio-ORACLE: A global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 272–281. <https://doi.org/10.1111/j.1466-8238.2011.00656.x>
- Tzadik, O.E., Jones, D.L., Peebles, E.B., Koenig, C.C., Stallings, C.D., 2017. The Effects of Spatial Scale on Assigning Nursery Habitats in Atlantic Goliath Groupers (*Epinephelus itajara*) Using Non-lethal Analyses of Fin Rays. *Estuaries and Coasts* 40, 1785–1794. <https://doi.org/10.1007/s12237-017-0244-z>
- Valentine, J.F., Heck, K.L., 2005. Perspective review of the impacts of overfishing on coral reef food web linkages. *Coral Reefs* 24, 209–213. <https://doi.org/10.1007/s00338-004-0468-9>

- van Proosdij ASJ, Sosef MSM, Wieringa JJ, Raes N. Minimum required number of specimen records to develop accurate species distribution models. *Ecography (Cop)*. 2016; 39: 542±552.
- Varela S, Anderson RP, García-Valdés R, Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography (Cop)*. Blackwell Publishing Ltd; 2014; 37: 1084±1091.
- Velazco S. J. E., Galvão, F., Villalobos, F., De Marco, P. 2017. Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. *PLoS ONE*, 12(10): e0186025.
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., & Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodiversity and Conservation*, 26(2), 251-271.
- Vollmar, A., Macklin, J.A., Ford, L., 2010. Natural History Specimen Digitization: Challenges and Concerns. *Biodivers. Informatics* 7, 93–112. <https://doi.org/10.17161/bi.v7i2.3992>
- Ward, R.D., Friess, D.A., Day, R.H., Mackenzie, R.A., 2016. Impacts of climate change on mangrove ecosystems: a region by region overview. *Ecosyst. Heal. Sustain.* 2. <https://doi.org/10.1002/ehs2.1211>
- Warren, D.L.; Glor, R.E.; Turelli, M.; Funk, D. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 2008, 62, 2868–2883.
- Watson, S.-A., Southgate, P.C., Tyler, P.A., Peck, L.S., 2009. Early Larval Development of the Sydney Rock Oyster *Saccostrea glomerata* Under Near-Future Predictions of CO₂-Driven Ocean Acidification . *J. Shellfish Res.* 28, 431–437. <https://doi.org/10.2983/035.028.0302>
- Weatherley, A., 1966. *Ecology Of Fish Growth*.
- Wernberg, T., Smale, D.A., Tuya, F., Thomsen, M.S., Langlois, T.J., De Bettignies, T., Bennett, S., Rousseaux, C.S., 2013. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat. Clim. Chang.* 3, 78–82. <https://doi.org/10.1038/nclimate1627>
- Werneck, F.P., Nogueira, C., Colli, G.R., Sites, J.W., Costa, G.C., 2012. Climatic stability in the Brazilian Cerrado: Implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *J. Biogeogr.* 39, 1695–1706. <https://doi.org/10.1111/j.1365-2699.2012.02715.x>
- Wieczorek, J., Guo, Q., Hijmans, R.J., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.* 18, 745–767. <https://doi.org/10.1080/13658810412331280211>
- Wilson, S.K., Burgess, S.C., Cheal, A.J., Emslie, M., Fisher, R., Miller, I., Polunin, N.V.C., Sweatman, H.P.A., 2008. Habitat utilization by coral reef fish: implications for

- specialists vs. generalists in a changing environment. *J. Anim. Ecol.* 281–291. <https://doi.org/10.1111/j.1365-2656.2007.0>
- Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A. Effects of sample size on the performance of species distribution models. *Divers Distrib.* 2008; 14: 763–773.
- Woodward, F. I. 1987. *Climate and plant distribution*. Cambridge: Cambridge University Press.
- Wunderlich RF, Lin Y-P, Anthony J, Petway JR. 2019. Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation* 35: 97–116.
- Yang, W., Ma, K., Kreft, H., 2013. Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *J. Biogeogr.* 40, 1415–1426. <https://doi.org/10.1111/jbi.12108>
- Zapelini, C., Giglio, V.J., Carvalho, R.C., Bender, M.G., Gerhardinger, L.C., 2017. Assessing Fishing Experts' Knowledge to Improve Conservation Strategies for an Endangered Grouper in the Southwestern Atlantic. *J. Ethnobiol.* 37, 478–493. <https://doi.org/10.2993/0278-0771-37.3.478>
- Zektser, I.S., Loaiciga, H.A., 1993. Groundwater fluxes in the global hydrologic cycle: past, present and future. *J. Hydrol.* 144, 405–427. [https://doi.org/10.1016/0022-1694\(93\)90182-9](https://doi.org/10.1016/0022-1694(93)90182-9)
- Zhang, L., Liu, S., Sun, P., Wang, T., Wang, G., Zhang, X., Wang, L., 2015. Consensus forecasting of species distributions: The effects of niche model performance and niche properties. *PLoS One* 10, 1–18. <https://doi.org/10.1371/journal.pone.0120056>
- Zhang, Z., Xu, S., Capinha, C., Weterings, R., Gao, T., 2019. Using species distribution model to predict the impact of climate change on the potential distribution of Japanese whiting *Sillago japonica*. *Ecol. Indic.* 104, 333–340. <https://doi.org/10.1016/j.ecolind.2019.05.023>
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Ritter, C.D., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengstrom, N., Zizka, V., Antonelli, A., 2019. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* 10 (5), 744e751.
- Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. *Ecography*, 43: 1261–1277.

Appendices

1. Marine ecoregions and model calibration area

The calibration area used in ENM/SDM models for *Epinephelus itajara* was delimited based on the World Marine Ecoregions (Spalding et al., 2007) in which species was present, plus those ecoregions immediately adjacent (Figure S1) (Table S1).

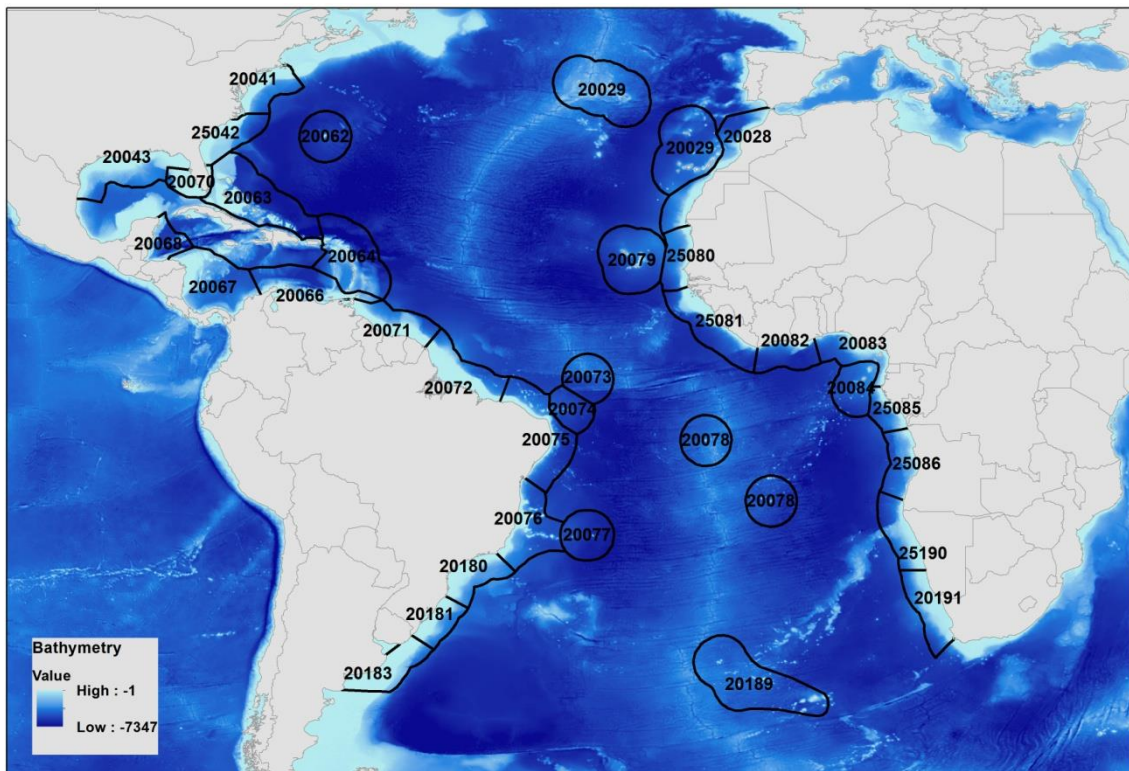


Figure S1. Ecoregions in the calibration area used in ENM/SDM models for *Epinephelus itajara*.

Table S1. Ecoregions in the calibration area used in ENM/SDM models for *Epinephelus itajara*.

ECO_CODE	ECOREGION
20072	Amazonia
20029	Azores Canaries Madeira
20063	Bahamian
20062	Bermuda
20079	Cape Verde
25042	Carolinian
20076	Eastern Brazil
20064	Eastern Caribbean
20074	Fernando de Naronha and Atoll das Rocas

20070	Floridian
20071	Guianan
20083	Gulf of Guinea Central
20084	Gulf of Guinea Islands
20082	Gulf of Guinea Upwelling
25081	Gulf of Guinea West
20191	Namaqua
25190	Namib
20075	Northeastern Brazil
20043	Northern Gulf of Mexico
20181	Rio Grande
20028	Saharan Upwelling
20073	Sao Pedro and Sao Paulo Islands
20180	Southeastern Brazil
20066	Southern Caribbean
20067	Southwestern Caribbean
20078	St. Helena and Ascension Islands
20077	Trindade and Martin Vaz Islands
20189	Tristan Gough
20183	Uruguay-Buenos Aires Shelf
20041	Virginian
20068	Western Caribbean
25085	Gulf of Guinea South
25086	Angolan
25080	Sahelian Upwelling

2. Jackknife test for environmental variables importance

The jackknife test was performed for 165 models: 11 variables * 5 algorithms (BRT, GLM, Maxent, RBF, SVM) * 3 types of occurrence dataset (Rd, Pd, FullSet) * one replica. We evaluated the value of AUC for each model created with only one variable and with all the remaining variables minus this one variable. These values indicated which variables contributed most for the model performance when used alone and which, when removed, impaired most the model performance (Table S2). After this test, we created a Spearman correlation matrix to analyze the correlation between the variables. At the end of this process we retained only the variables that matter most to the model and the lowest correlated ($r^2 < |0.7|$).

Table S2: Jackknife test results showing the performance (in AUC values) of the ENM/SDM models generated for *Epinephelus itajara* with only and without each environmental variable. Biogeo08: Mean Annual Salinity, biogeo09: Salinity of the Freshest Month, biogeo10: salinity of the saltiest month, biogeo11: Annual range in salinity, biogeo12: annual variance in salinity, biogeo13: Mean annual in temperature, biogeo14: temperature of the coldest month, biogeo15: temperature of the warmest month, biogeo16: annual range in temperature, biogeo17: annual variance in temperature.

Variables	With Only	Without
Bathymetry	0.804	0.155
Biogeo08	0.303	0.215
Biogeo09	0.238	0.185
Biogeo10	0.184	0.198
Biogeo11	0.364	0.205
Biogeo12	0.412	0.223
Biogeo13	0.265	0.194
Biogeo14	0.304	0.165
Biogeo15	0.299	0.179
Biogeo16	0.04	0.200
Biogeo17	0.189	0.190

3. Pseudo-absence generation

Given that, except Maxent, the other four chosen modeling algorithms work with presence-absence (or pseudo-absence) data, we tested the effect of different amounts of pseudo-absence points. The number of pseudo-absences can influence models accuracy (ability to predict suitable or unsuitable areas). Algorithms of regression techniques seem to work best with a high number of pseudo-absences, whereas classification and machine learning algorithms tend to have a better result with a low or moderate amount of pseudo-absences, i.e., number of pseudo-absences equalized to the presences or a little higher (Barbet-Massin et al., 2012). Therefore, we performed a sensitivity analysis creating models with different proportions of pseudo-absences in relation to the set of presences (1:1, 2:1, 10:1 and 100:1). Models generated with a proportion of 1:1 obtained, on average, the highest sensitivity value (0.95 ± 0.06) (Table S3). Thus, we generate pseudo-absences randomly in the geographic calibration space, excluding cells with detected presence plus a 50km buffer, respecting the proportion of one pseudo-absence for each presence record.

Table S3: Sensibility analysis results showing the performance (in Sensitivity values) of the ENM/SDM models generated for *Epinephelus itajara* with different presence/absence ratios for selection of pseudo-absences.

Presence/absence ratio	Sensitivity
1/1	0.950 ± 0.06
2/1	0.835 ± 0.13
10/1	0.809 ± 0.23
100/1	0.767 ± 0.27

4. Environmental suitability predictions

The follow maps show the geographic distribution of the environmental suitability generated by the ensemble models for the three occurrence datasets FullSet (Figure S2), RdPd (Figure S3), and PdRd (Figure S4).

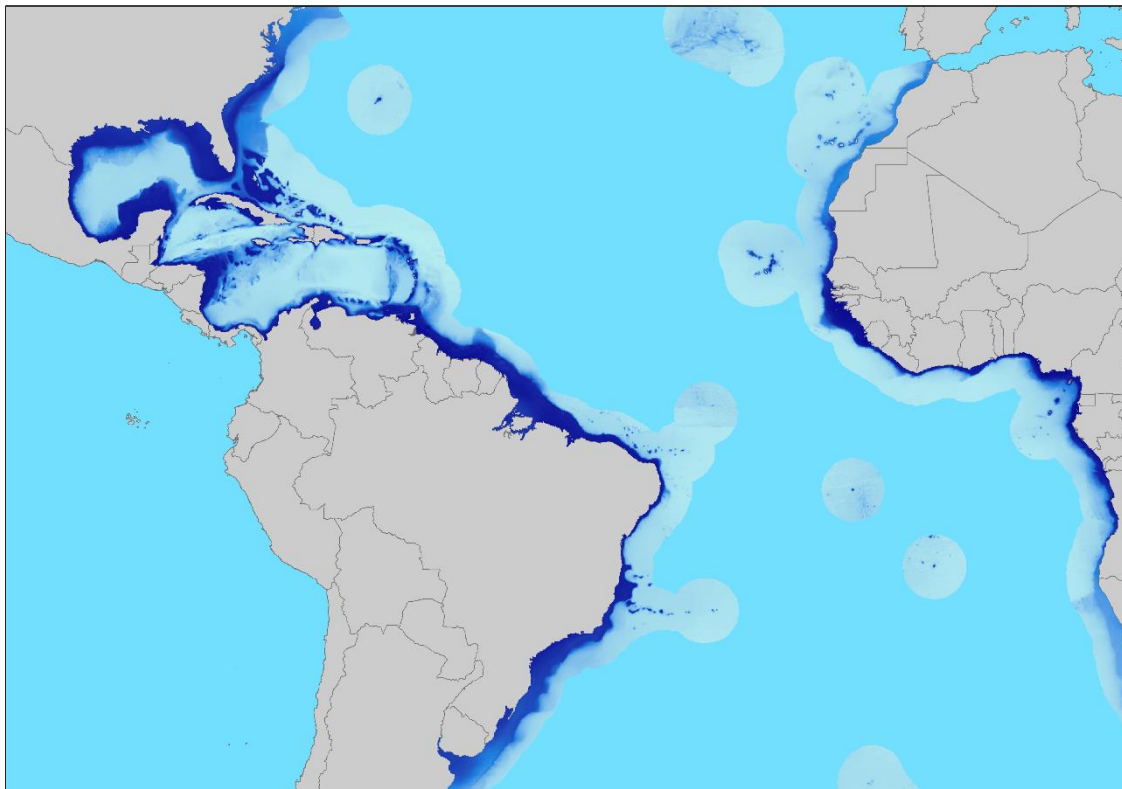


Figure S2: Ensemble model for FullSet occurrence dataset. Darker blue color are high suitability cells and light blue colors are low suitability cells. Values range from 0.02 to 0.94.

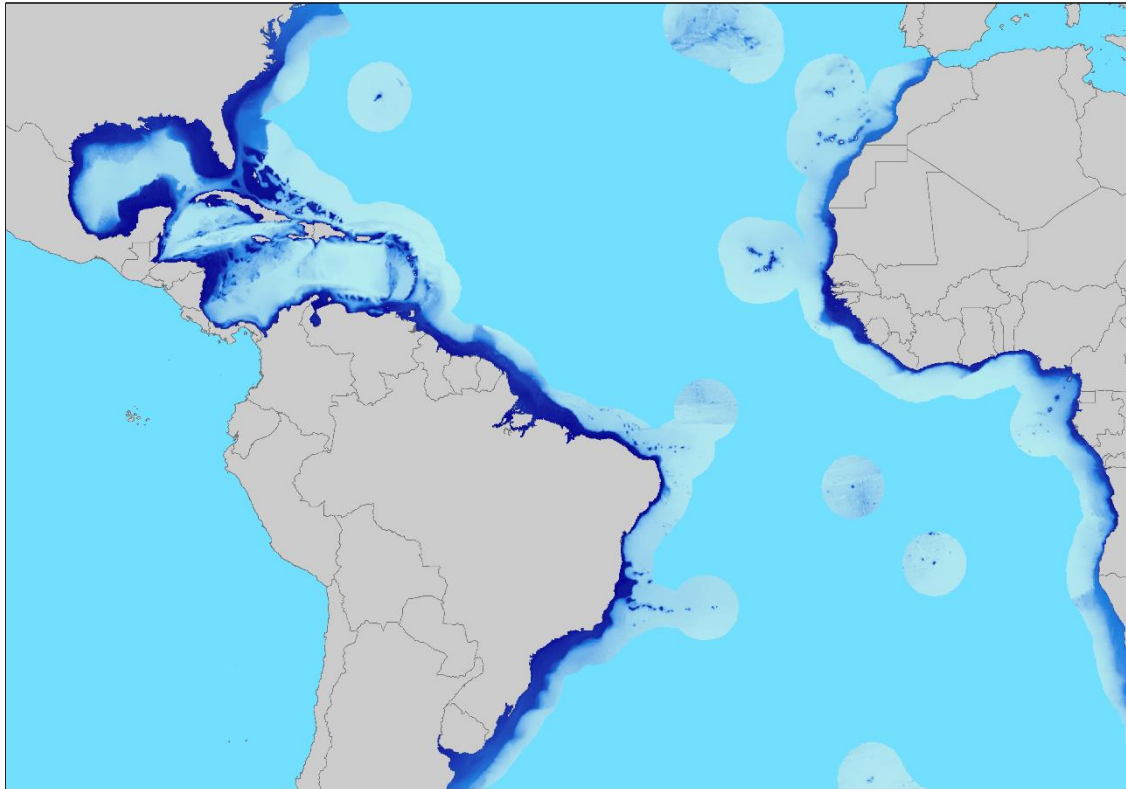


Figure S3: Ensemble Model for RdPd occurrence dataset. Darker blue colors are high suitability cells and light blue colors are low suitability cells. Values range from 0.07 to 0.95.

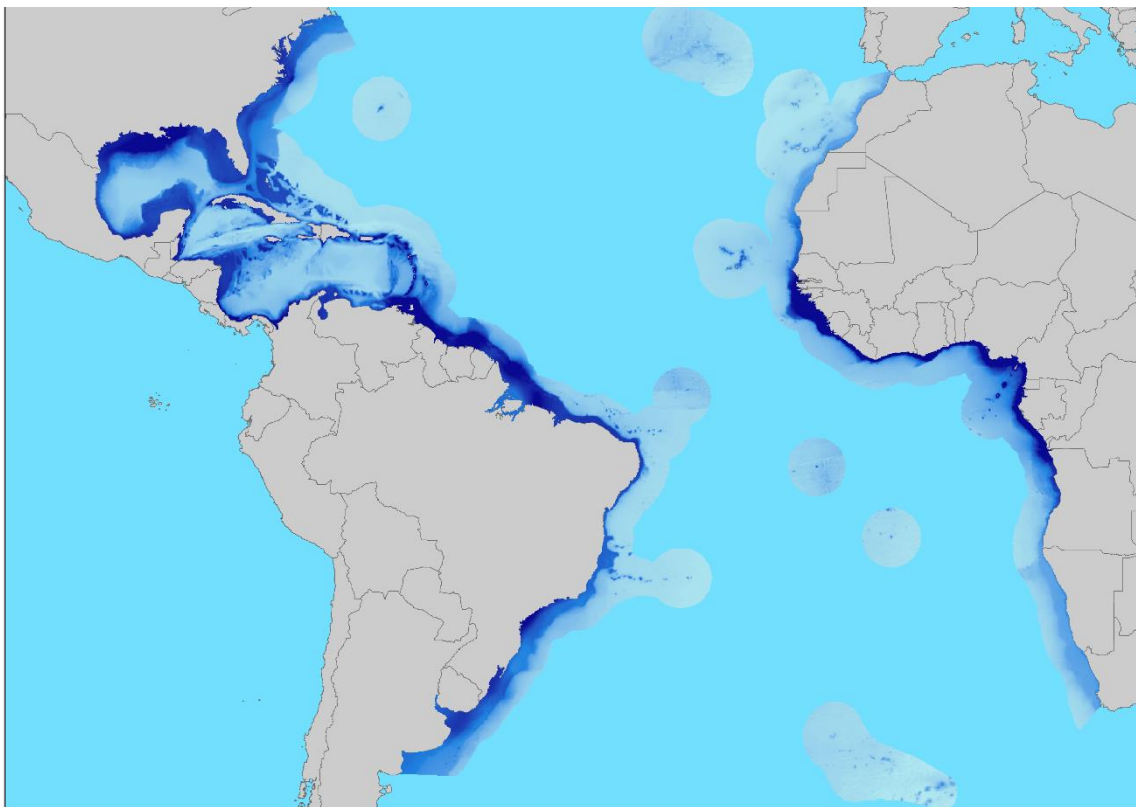


Figure S4: Ensemble Model for PdRd occurrence dataset. Darker blue colors are high suitability cells and light blue colors are low suitability cells. Values range from 0.00 to 0.95.

5. Selection of variables for the Goliath grouper ecological niche model

The Jackknife test is a resampling method that consists of removing an element from the original set and observing the effect of the remaining elements, repeating it n times (where n is the number of element). Therefore, we created ecological niche models using 11 environmental variables minus the retained variable, repeating the process until all variables have been retained ($n = 12$), being able, then, to observe how much the model had gains, through the Area Under the ROC Curve (AUC), when we removed the variable in question. Furthermore, we did the opposite process, in which we built ecological niche models only with the variable retained in the previous step, to analyze how good the model is only with that variable, repeating until models with all the variables have been made (Table S4). From this analysis, we created a ranking of the most important variables and compared how correlated they were with the rest of the group, through Spearman's correlation matrix (Figure S5). In this way, we eliminate the variables that are most correlated with those that are in the first positions of the ranking, retaining, at the end of the process, the most important and least correlated variables. This procedure was done using the `getVarImp` function in the SDM package on R.

Table S4: Table with the contribution values of the variables for the niche models. In the Without column are the contribution values for the model when the variable in question is removed from the model construction and in the With Only column are the contribution values for the model when only the variable in question is used for the model construction. Salinity.Lt.max = average of the maximum salinity records, Salinity.Lt.min = average of the minimum salinity records, Salinity.max = maximum salinity value, Salinity.mean = average of the salinity values, Salinity.min = value minimum salinity, Salinity.range = range of salinity values, Temperature.Lt.max = average of maximum temperature records, Temperature.Lt.min = average of minimum temperature records, Temperature.max = maximum temperature value, Temperature.mean = average of temperature values, Temperature.min = minimum temperature value, Temperature.range = range of temperature values.

Variable	Without	With Only
Salinity.Lt.max	0.067767	0.05144
Salinity.Lt.min	0.180676	0.50192
Salinity.max	0.130429	0.03752
Salinity.mean	0.160065	0.30728
Salinity.min	0.055564	0.50304
Salinity.range	0.044276	0.03672
Temperature.Lt.max	0.07184	0.24176
Temperature.Lt.min	0.080604	0.16296
Temperature.max	0.071716	0.24792
Temperature.mean	0.114836	0.32592
Temperature.min	0.110436	0.25488
Temperature.range	0.124371	0.0008

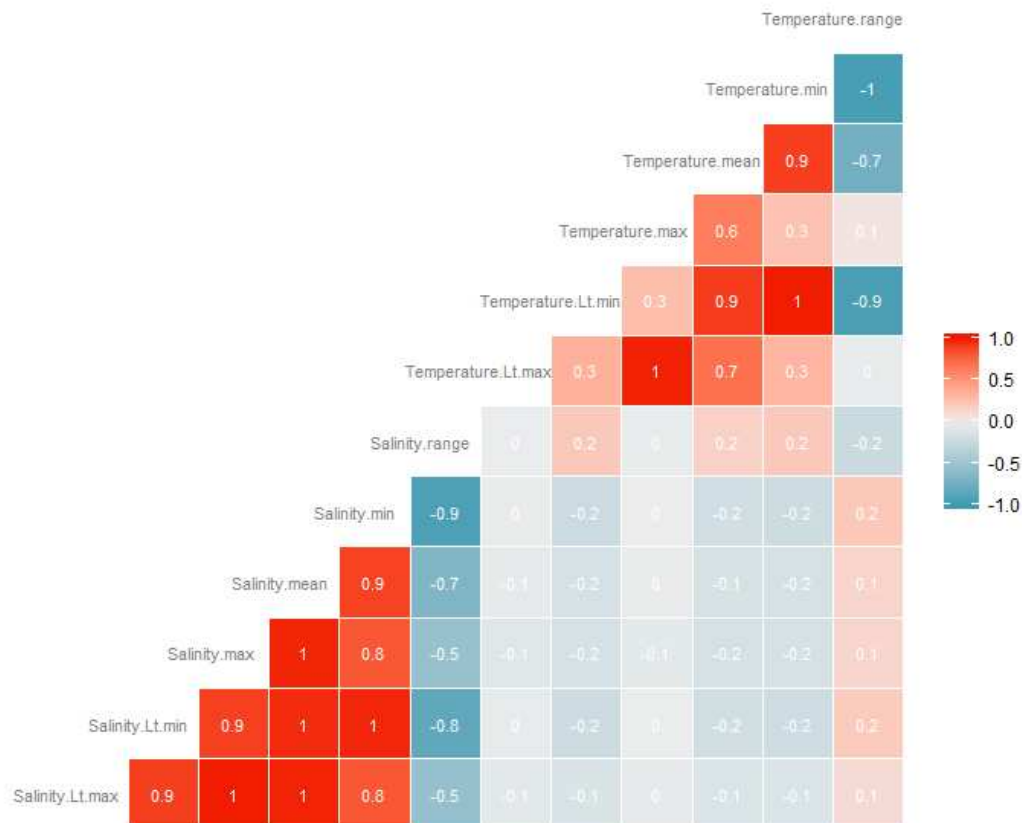


Figure S5: Spearman's correlation matrix with the correlation values for the pairs of variables. Reddish colors indicate strong positive correlations and blueish colors indicate strong negative correlations.

6. Development of habitat predictors

To develop habitat predictors we first applied an ENM/SDM approach to generate environmental suitability for each coastal habitat type (seagrass, mangrove, rocky shores and coral reefs), and next we generated rasters representing distance from the suitable area for these habitats.

6.1. Occurrence records for habitats

Information on seagrass distribution derived from the World Atlas of Seagrass. This dataset represents the compilation of occurrence polygon maps and points of seagrass around the world. Due to the heterogeneity of the data and sampling problems, we only use information from polygons. The geographic distribution of mangroves derived from the occurrence polygon maps provided by USGS, generated from Global Land Surveys images at a resolution of 30 meters (Global Mangrove Distribution - USGS). Following Halpern et al. (2008), for rocky shores we assume a distribution encompassing any cell

distant 1km from the coast line. Therefore, from a detailed coastline provided by OpenStreetMap (<https://osmdata.openstreetmap.de/>), we created a 1 km buffered polygon over the entire model calibration area to represent the geographic extent of rocky reef. Information about coral reef distribution was extracted from the Reef at Risk Revisited, which combines a large number of coral reef maps derived from different sources.

To obtain occurrence records for the ecological niche modeling, we randomly generated georeferenced points distant 1 km from each other inside polygons of each habitat type, and then we apply a spatial rarefaction to eliminate points nearer than 10 km in order to avoid spatial autocorrelation effects that could compromise modeling results (F. Dormann et al., 2007). Table S5 shows the total occurrence records obtained for each habitat type. All of these procedures were performed with ArcGIS and SDMToolBox tool.

6.2. Variables selection for ecological niche and environmental suitability modeling

To select environmental variables for ecological niche models of the habitats we applied a PCA approach, analysing the first two axes of the main components of the entire set of environmental variables. For seagrass, rocky shores and coral reefs we use only marine variables from BioOracle. For mangroves which are influenced by both marine (e.g. sea temperature and salinity) and atmospheric (temperature and precipitation) components (Ward et al., 2016), we used bioclimatic variables from WorldClim in addition to BioOracle variables in order to more accurately capture the environmental suitability for this system (Table S5).

Table S5: Number of georeferenced and unique occurrence records and environmental variables used for ecological niche modeling of habitats. Salinity.Lt.max = average of maximum salinity records,

Salinity.Lt.min = average of minimum salinity records, Salinity.range = range of salinity values, Temperature.Lt.max = average of maximum temperature records, Temperature.Lt.min = average of the minimum temperature records, bio 05 = Maximum temperature of the hottest months, bio06 = Minimum temperature of the coldest months, bio13 = Precipitation of the wettest months, bio14 = Precipitation of the driest months.

Habitat type	Number of records	Environmental variables
Seagrass	157	Salinity.Lt.max, Temperature.Lt.max, Temperature.Lt.min
Mangrove	3923	Salinity.Lt.min, Temperature.Lt.max, bio05,bio06,bio13,bio14
Rocky reef	2675	Salinity.Lt.max, Salinity.range, Temperature.Lt.max, Temperature.Lt.min
Coral reef	311	Salinity.Lt.max, Salinity.Lt.min, Temperature.Lt.max, Temperature.Lt.min

6.3. Ecological niche and environmental suitability modeling for habitats and derived distance variables

To generate environmental suitability we ran 500 replicas of ecological niche models (5 algorithms * 100 replicates per algorithms) for each habitat, using the bootstrapping method implemented in the SDM package on R platform. From the confusion matrix we calculated model performance metrics of sensitivity, specificity and True Skill Statistics - TSS (Allouche et al., 2006), based on the threshold rule that maximizes the sum of sensitivity with specificity (Liu et al., 2013). We only use models that have at least 0.700 sensitivity and specificity to generate binary consensus maps of environmental suitability (Figures S6 to S9). From these binary maps, in ArcMap software we created a continuous raster representing the Euclidean distance from each cell of suitable area (cell value = 1) (Chapter 2, Figures 2 to 5). Using the map algebra in ArcMap, we summed the values of the same pixel for all habitats in order to obtain a raster representing the Euclidean distance for any type of habitat (Chapter 2, Figure 6). These Euclidean distance rasters were used as habitat variables in our ecological niche models for *Epinephelus itajara*.



Figure S6: Binary suitability map for seagrass in the Atlantic Ocean. Dark blue are the areas of presence (cell value = 1) of suitable area. Light blue are not suitable areas (cell value = 0). Green dots are the 157 unique records used for building the ecological niche model for this habitat.

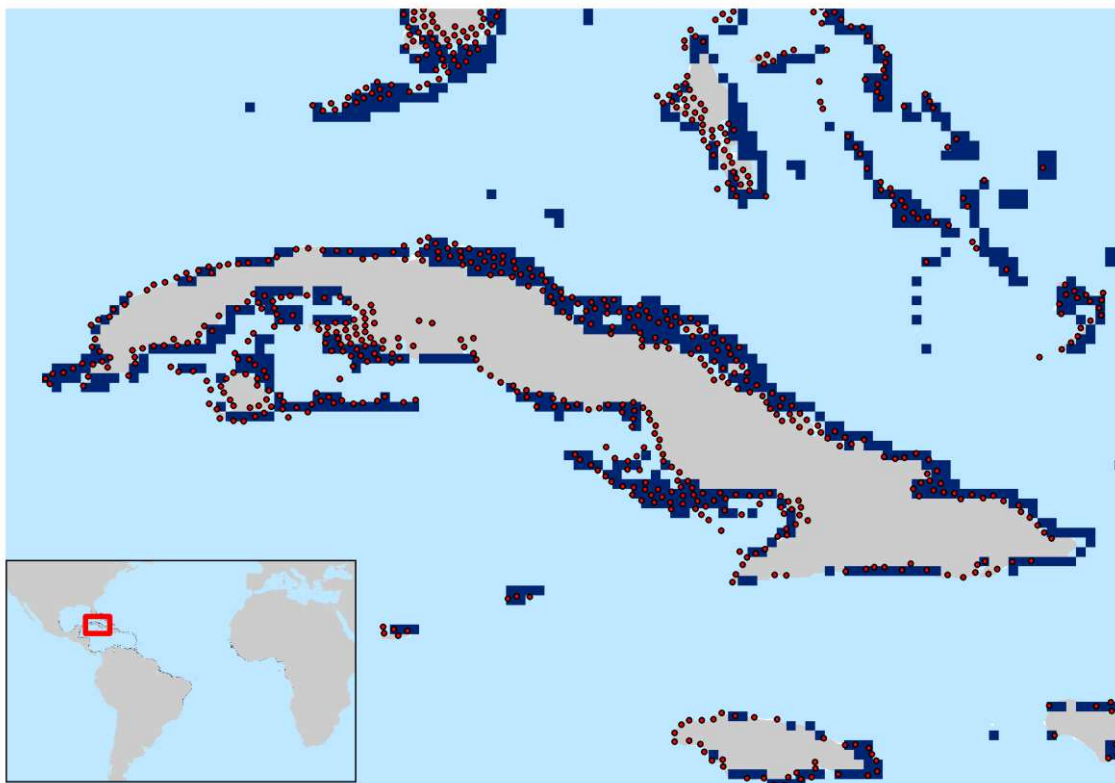


Figure S7: Binary suitability map for mangrove in the Atlantic Ocean. In detail the island Cuba where dark blue are the areas of presence (cell value = 1) of suitable area. Light blue are not suitable areas (cell value = 0). Red dots are the 3923 unique records used for building the ecological niche model for this habitat.

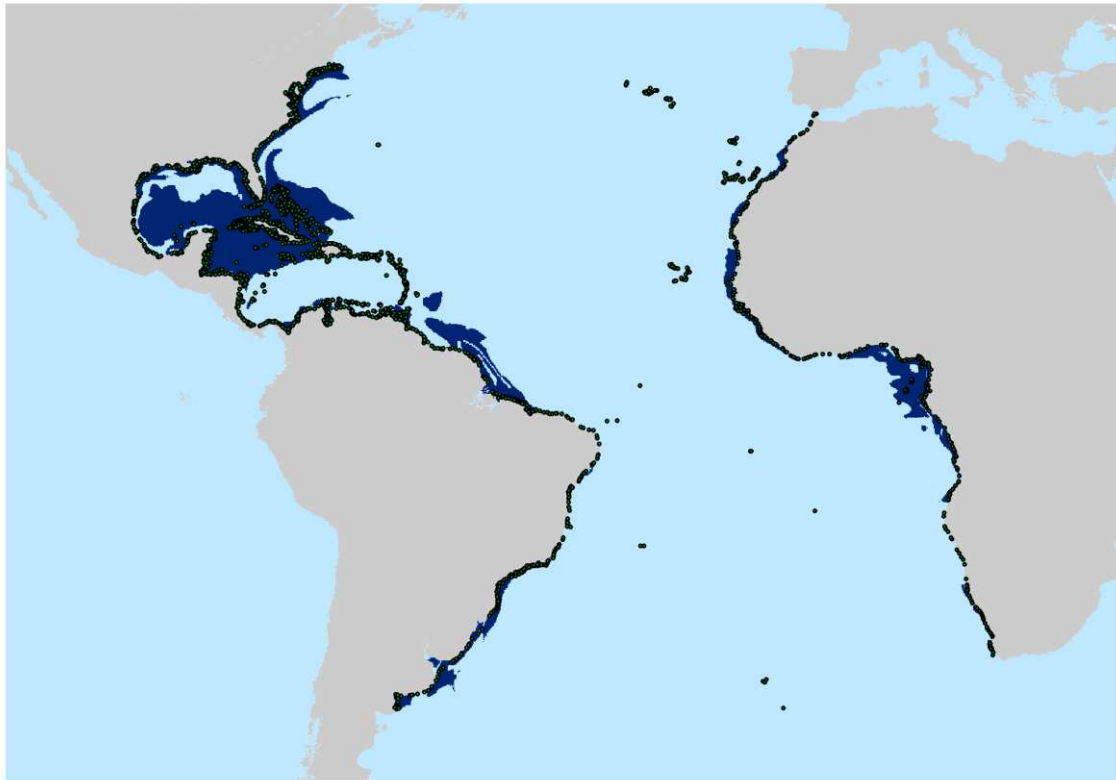


Figure S8: Binary adequacy map for rockyreef in the Atlantic Ocean. Dark blue are the areas of presence (cell value = 1) of suitable area. Light blue are not suitable areas (cell value = 0). Green points are the 2675 unique records used for the construction of the ecological niche model for this habitat.



Figure S9: Binary adequacy map for coralreef in the Atlantic Ocean. Dark blue are the areas of presence (cell value = 1) of suitable area. Light blue are not suitable areas (cell value = 0). Green dots are the 311 unique records for building the ecological niche model for this habitat.

7. Pseudo-absences and data partitioning settings for the Goliath grouper ecological niche model

The methods chosen for pseudo-absence generation and data partitioning affected output and performance of ENMs (Morgane Barbet-Massin et al., 2012; Radosavljevic & Anderson, 2014; Iturbide et al., 2015; Roberts et al., 2017). In this regard, we performed a sensitivity analysis to define better parameterization for Goliath grouper ecological niche models. To conduct a species-specific tuning of model settings, we generated models by the methods of cross-validation (with 5 and 10 folds) and subsampling (90% of the records for training and 10% for validation). For each partitioning method we tested different pseudo-absences to presences ratios (1:1, 2:1, 10:1). The choice of the best model was based on the sensitivity metric, a statistic that measures the omission error (number of presences incorrectly predicted as absence) in the geographic space. Our results indicated subsampling method with a pseudo-absence ratio for the presence of 1/1 as best parameters (Table S6).

Tabela S6: Results of the sensitivity analysis for model parameterization of data partitioning and pseudo-absences to presences ratio. 1xCV10 = cross-validation with 10 folds and 1:1 pseudo-absence to presences ratio, 2xCV10 = cross-validation with 10 folds and 2:1 pseudo-absence to presences ratio, 10xCV10 = cross-validation with 10 folds and 10:1 pseudo-absence to presences ratio, 1xCV5 = cross-validation with 5 folds and 1:1 pseudo-absence to presences ratio, 2xCV5 = cross-validation with 5 folds and 2:1 pseudo-absence to presences ratio, 10xCV5 = cross-validation with 5 folds and 10:1 pseudo-absence to presences ratio, 1xSub = subsampling with 90% of training points and 10% for validation and 1:1 pseudo-absence to presences ratio, 2xSub = subsampling with 90% of points for training and 10% for validation and 2:1 pseudo-absence to presences ratio, 10xSub = subsampling with 90% of points for training and 10% for validation and 10:1 pseudo-absence to presences ratio. Best parameters and correspondent sensitivity value are shown in boldface.

Parameters	1xCV10	2xCV10	10xCV10	1xCV5	2xCV5	10xCV5	1xSub	2xSub	10xSub
Sensitivity	0,762	0,734	0,713	0,737	0,757	0,738	0,769	0,709	0,726

8. Anthropogenic stressors and Marine Protected Areas data

Data about anthropogenic stressors were obtained from Halpern (2015) and included variables that summarize the impacts of the main threats for *Epinephelus itajara* according IUCN assessments (Craig et al., 2009; Bertonecini et al., 2018). The four normalized variables used were: (i) artisanal fishing (Figure S10), (ii) demersal, non-destructive, lowbycatch (Figure S11), (iii) ocean pollution (Figure S12), and (iv) population pressure (Figure S13) (see Halpern, 2015, supplementary material, for

details about how these variables were created). In addition to these four variables, we used map algebra to generate a cumulative impact variable by summing all the impact rasters (Figure S14).

Marine Protected Areas (MPAs, IUCN categories I-VI) came from the World Database of Protected Areas (WDPA, <http://protectedplanet.net>) (Figure S15).

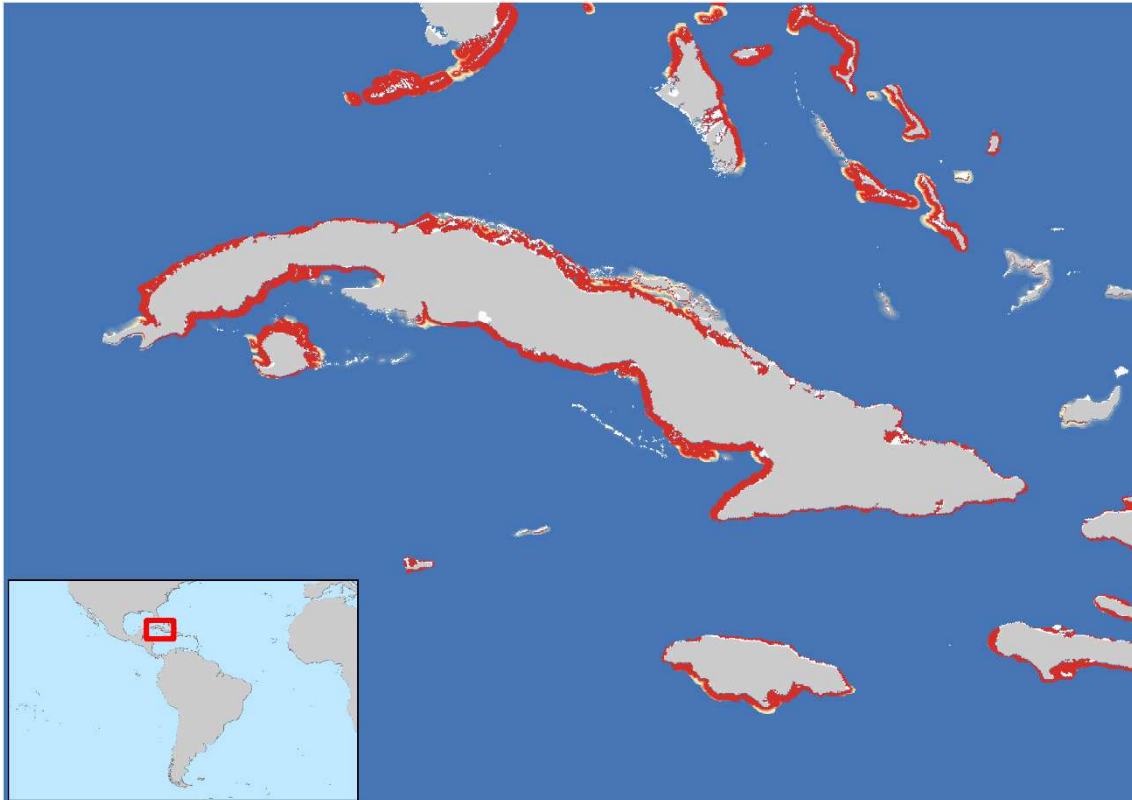


Figure S10: Artisanal fishing variable. In detail, the island of Cuba where reddish colors are areas where there is high pressure of artisanal fishing and lighter colors where there is low presence of artisanal fishing.

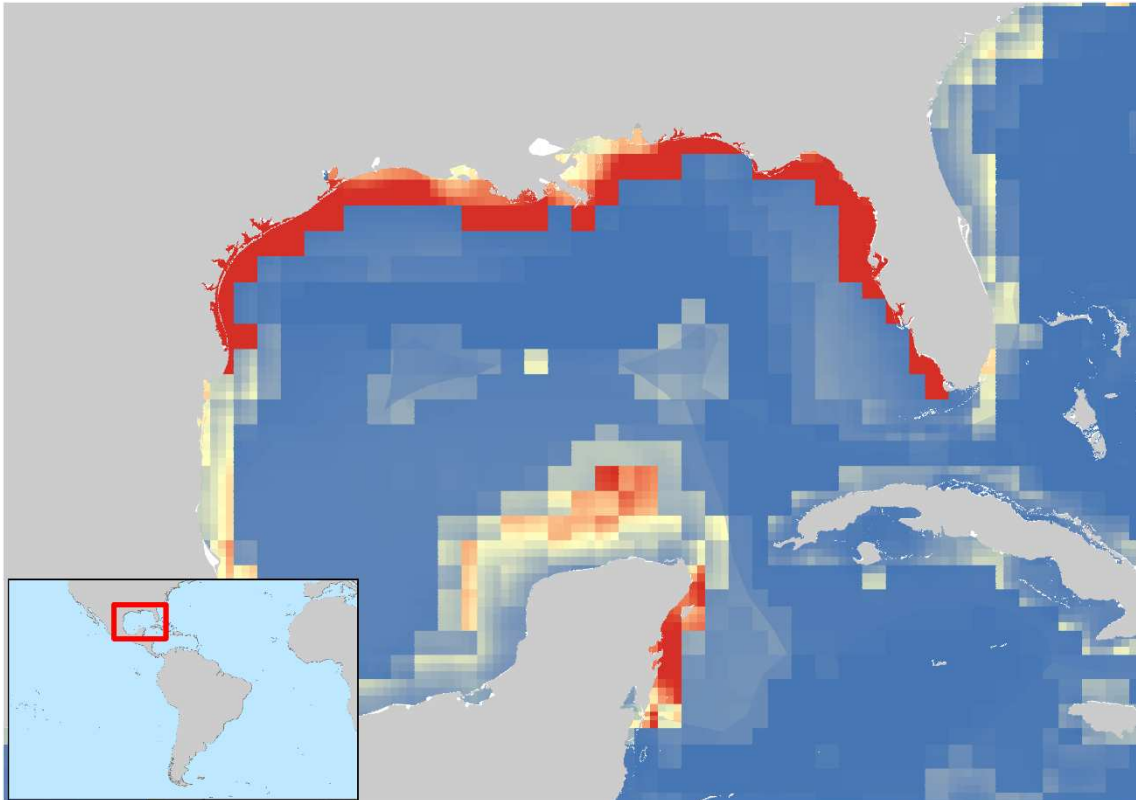


Figure S11: demersal, non-destructive, lowbycatch variable. In detail Gulf of Mexico where reddish colors are areas where there is a high presence of demersal, non-destructive, lowbycatch and lighter colors are areas where there is a low presence of demersal, non-destructive, lowbycatch.

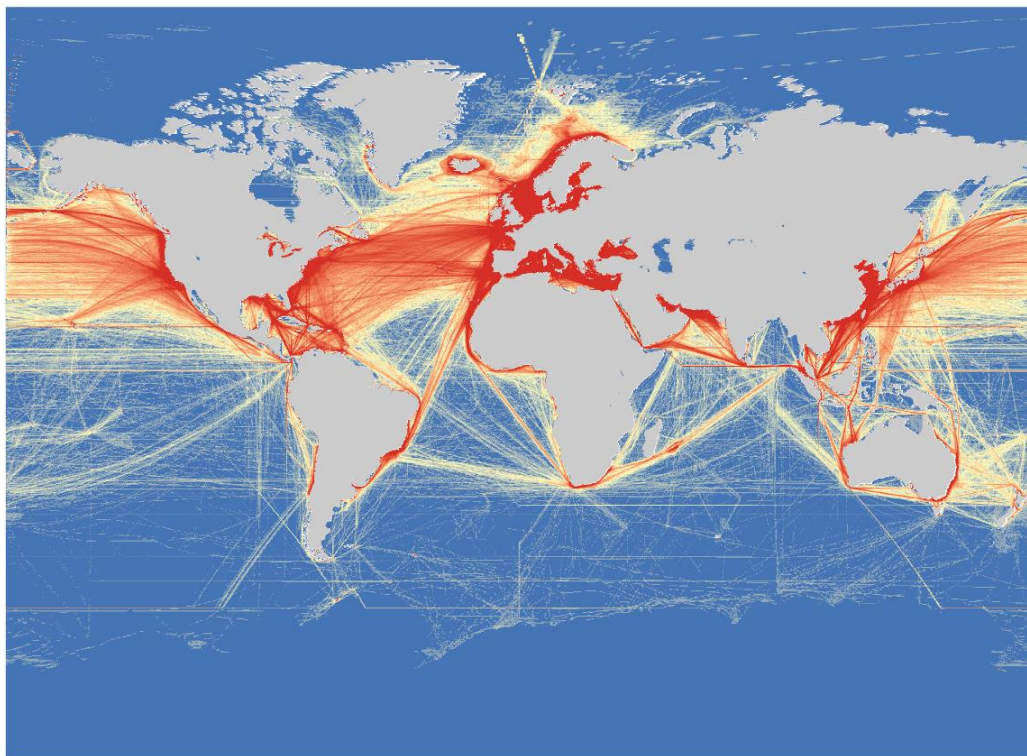


Figure S12: Ocean pollution variable. Reddish colors are areas where there is a high concentration of ocean pollution and lighter colors are areas where there is a low concentration of ocean pollution.

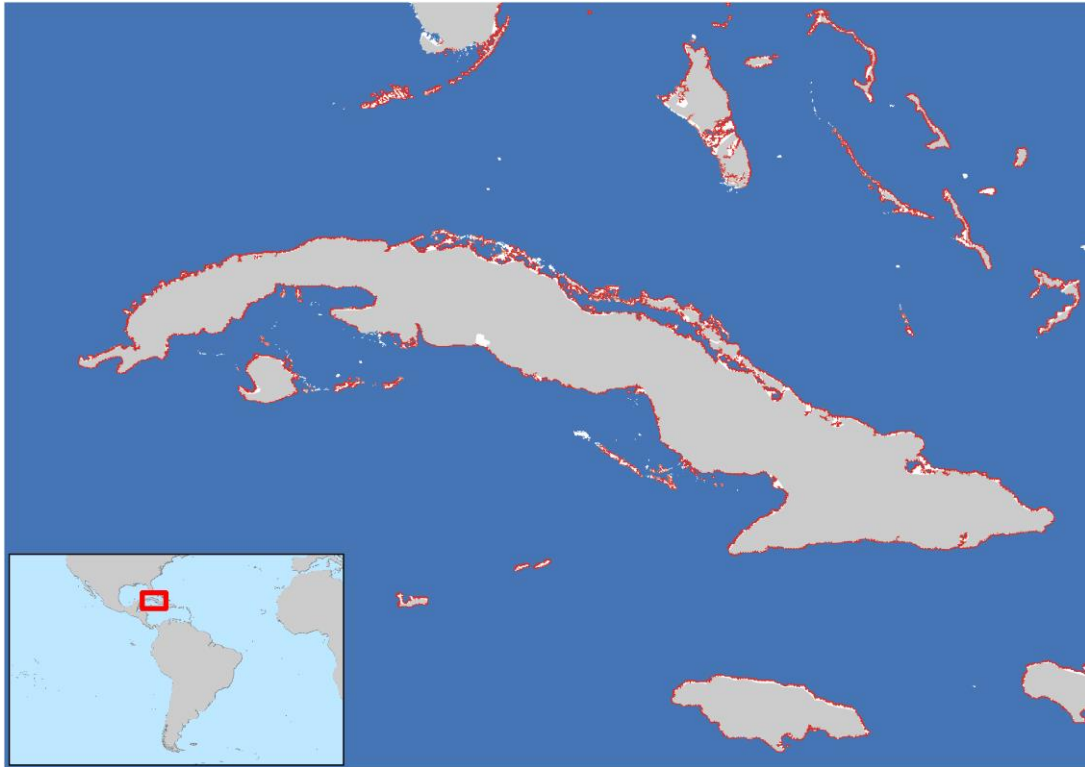


Figure S13: Population pressure variable. In detail, the island of Cuba where reddish colors are areas where there is a high pressure of population pressure and lighter colors where there is a low presence of population pressure.

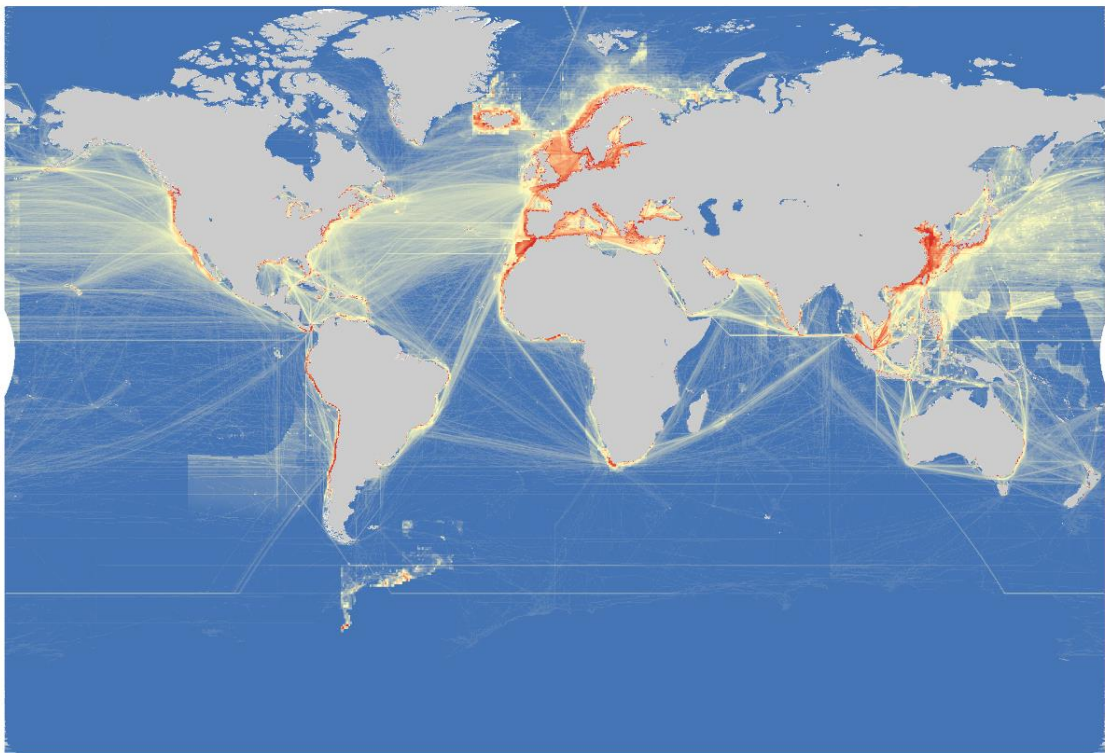


Figure S14: Cumulative impact variable. Reddish colors are areas where there is a high impact of the four variables together and lighter colors are areas where there is a low impact of the four variables together.

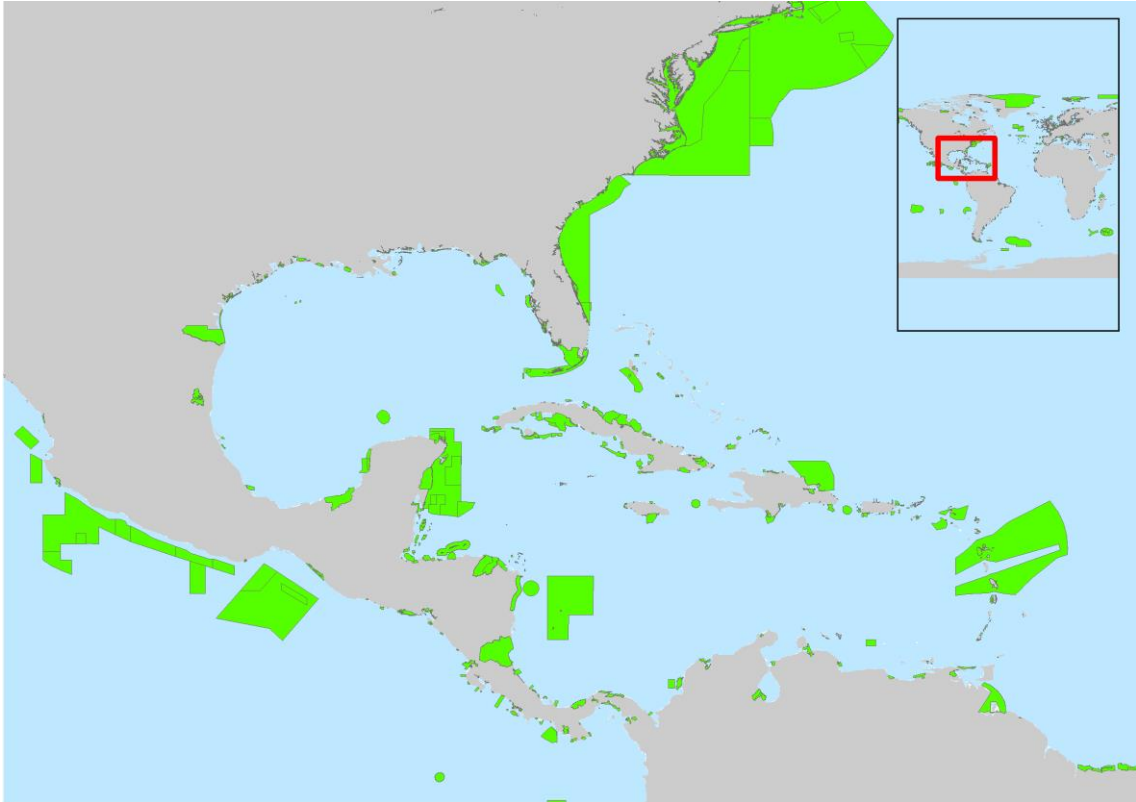


Figure S15: Map of Marine Protected Areas (MPAs). In detail, the Gulf of Mexico and the Caribbean Sea where MPAs areas are represented by green polygons.