Ariadni Vafeiadou

Global biogeographic patterns of marine sponges

2023

Ariadni Vafeiadou

Global biogeographic patterns of marine sponges

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2023

Ariadni Vafeiadou

Global biogeographic patterns of marine sponges

Mestrado em Biologia Marinha

Supervisors:
Jorge Assis

UNIVERSIDADE DO ALGARVE

UNIVERSIDADE DO ALGARVE

Faculdade de Ciências e Tecnologia

2023

Declaração de autoria de trabalho

Global biogeographic patterns of marine sponges

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Signature

Ariadni Vafeiadou

Declaração de autoria de trabalho

A Universidade do Algarve reserva para si o direito, em conformidade com o disposto no Código do Direito de Autor e dos Direitos Conexos, de arquivar, reproduzir e publicar a obra, independentemente do meio utilizado, bem como de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição para fins meramente educacionais ou de investigação e não comerciais, conquanto seja dado o devido crédito ao autor e editor respetivos.

Signature

Ariadni Vafeia

Resumo

Na era do Antropoceno, as atividades antropogénicas tiveram um impacto dramático no planeta, levando a alterações significativas nos seus elementos, incluindo o oceano global. As maiores ameaças aos ambientes marinhos incluem a sobrepesca, a perda de habitat, a poluição e fatores de stress induzidos pelas alterações climáticas, como a acidificação das águas e as temperaturas elevadas. Mais precisamente, para taxa bentónicos sésseis, em particular as esponjas - que desempenham papéis cruciais nos ecossistemas marinhos - a pesca de arrasto e as atividades petrolíferas e de gás estão entre as ameaças mais alarmantes. Estas perturbações mediadas pelo homem desencadearam uma perda de biodiversidade, afetando o funcionamento dos ecossistemas e, consequentemente, o bem-estar humano. Reconhecendo os efeitos destas atividades, os países uniram-se e tentaram combater coletivamente o declínio da biodiversidade global, como a Meta de Biodiversidade de 2010 e o Plano Estratégico para a Diversidade 2011-2020, mas não conseguiram atingir esses objetivos. Atualmente, está em vigor o Quadro Mundial para a Biodiversidade pós-2020, que visa igualmente uma mudança positiva em termos de biodiversidade. No centro da abordagem das questões de biodiversidade, a compreensão dos padrões globais de biodiversidade é extremamente importante, mas continua a ser limitada, especialmente nos ambientes marinhos. Por esta razão, iniciativas globais, como o Censo da Vida Marinha, procuram discernir estes padrões, reconhecendo a importância de estabelecer uma base de referência para monitorizar futuras alterações da biodiversidade. No entanto, para gerir a biodiversidade, é necessário compreender a forma como as espécies se distribuem no mar, sendo a riqueza e a endemicidade das espécies as principais métricas. Os Modelos de Distribuição de Espécies (MDE) são uma ferramenta fundamental que oferece informações valiosas sobre a distribuição das espécies, sendo utilizados em diversas aplicações em ecologia, evolução e ciências da conservação. Os MDE são ferramentas computacionais concebidas para quantificar a relação entre a distribuição de uma espécie e múltiplos fatores de previsão ambiental. Associam observações da ocorrência ou abundância de espécies a estimativas ambientais utilizando uma variedade de métodos estatísticos. A utilidade dos modelos depende da avaliação do seu desempenho, que identifica os pontos fortes e as limitações, determinando assim o âmbito da sua aplicação prática. Devido à crescente disponibilidade de métodos estatísticos e de dados biológicos e ambientais, a sua aplicação tem-se desenvolvido em ecossistemas terrestres, de água doce e marinhos. As esponjas são componentes essenciais das comunidades bentónicas, tendo colonizado todos os ecossistemas aquáticos. Desempenham uma multiplicidade de papéis funcionais no que respeita ao substrato, à ligação bento-pelágica e a associações com outros organismos. Avanços recentes conduziram a estimativas globais da diversidade marinha e do endemismo entre taxa, mas a maior parte da investigação sobre a distribuição centra-se nos peixes, mamíferos e moluscos. No entanto, as esponjas, apesar dos seus diversos papéis e da sua presença generalizada nos ecossistemas aquáticos, são frequentemente negligenciadas. Existe apenas um número limitado de publicações sobre a distribuição de esponjas, e estas centram-se em espécies ou extensões geográficas específicas e limitadas, ou em associação com outros taxa. Devido à preocupação global com a perda de biodiversidade marinha, há uma procura crescente de determinar os atuais padrões de distribuição de espécies desconhecidas para identificar e medir eficazmente as alterações na biodiversidade. Para responder a esta necessidade, este estudo procurou estimar os padrões de biodiversidade global das demosponjas marinhas, membros da classe Demospongiae, que compreende ~81% do filo total. Numa primeira fase, foram recolhidos registos de espécies desta classe em diversos repositórios online. As observações recolhidas foram submetidas a uma validação cruzada para detetar potenciais erros taxonómicos e de distribuição espacial. Os registos que não se enquadravam nas suas distribuições batimétricas ou geográficas conhecidas ou em terra, de acordo com o conhecimento de especialistas, foram removidos. Obteve-se assim um conjunto de dados de 376.242 registos validados de 1.808 espécies de demosponjas taxonomicamente aceites. Posteriormente, este conjunto de dados foi utilizado juntamente com um conjunto de variáveis ambientais ecologicamente relevantes para produzir modelos de distribuição de espécies. Foi utilizado um conjunto de três algoritmos: Boosted Regression Trees, Adaptive Boosting, Extreme Gradient Boosting. O método seguido produziu modelos individuais para 649 espécies de demosponge, que foram depois agrupados para revelar os seus padrões de biodiversidade a nível global. Foram utilizadas três métricas para avaliar o desempenho dos modelos Area Under the Curve, True Skill Statistic and Boyle's metric. Adicionalmente, a contribuição relativa global dos preditores ambientais foi estimada, juntamente com limites considerados críticos para a disatribuição das espécies. Além disso, para explorar melhor estes padrões, foi efetuada uma análise hierárquica de agrupamento, com base nos dados de presença/ausência das 649 espécies em 12 Eco-regiões Marinhas do Mundo. Este estudo fornece, pela primeira vez, uma visão global dos padrões de distribuição global das espécies de demosponge com recurso à modelação da distribuição das espécies, melhorando a nossa compreensão deste grupo. De acordo com as três métricas, o desempenho do modelo é excelente e a sua incerteza é baixa. Com base nestes resultados, as condições térmicas caracterizam predominantemente o habitat das demosponjas, mas a produtividade primária e o pH também desempenham um papel significativo na sua distribuição. A maior diversidade está prevista nas regiões tropicais e subtropicais. Nomeadamente, áreas como o Mar das Caraíbas e o Golfo do México no Oceano Atlântico, bem como o Triângulo de Coral e as áreas da Grande Barreira de Coral nos Oceanos Indo-Pacífico, apresentam uma elevada biodiversidade. A técnica de agrupamento revelou três grupos distintos de demosponjas. O Oceano Austral emergiu como um grupo único, refletindo elevados níveis de endemicidade. O Pacífico Tropical Oriental emparelhou com a América do Sul Temperada. O resto das regiões marinhas juntaram-se num terceiro grupo, indicando que o Atlântico, em conjunto com o Indo-Pacífico e a Australásia, albergam uma parte substancial de espécies com distribuições generalizadas. As estimativas providenciadas, que identificam áreas de elevada diversidade de demosponjas, podem agora informar estratégias de conservação e gestão, incluindo ações destinadas a salvaguardar as suas numerosas contribuições ecológicas.

palavras-chave: modelação da distribuição das espécies, padrões de biodiversidade, registos de ocorrência, riqueza das espécies, esponjas, endemismo

Abstract

Anthropogenic activities have dramatically impacted our oceans, which have triggered cascading biodiversity loss. This, in turn, affects ecosystem functioning, services, and ultimately human well-being. In response, international agreements set conservation goals for positive biodiversity. A prerequisite for effective management is a comprehensive understanding of current marine species distribution. Sponges are essential components of benthic communities, having colonized all aquatic ecosystems. Despite their diverse ecological roles, sponges are often overlooked. In this study, we focused on marine sponges from the class Demospongiae, which make up ~ 81% of all sponges. Using Species Distribution Models, we delineated global biodiversity patterns of marine demosponges. Occurrence records, taxonomically and spatially curated, were paired with relevant environmental predictors using an ensemble of machine learning algorithms. The resulting individual models were stacked to unravel biodiversity trends on a global scale, along with overall relative predictors' contribution and their corresponding critical thresholds. The findings showed that demosponge habitats are predominantly influenced by thermal conditions, pH, and phytoplankton productivity. Highest diversity was estimated in the tropical and subtropical zones, with one hotspot residing in the Caribbean Sea and Gulf of Mexico in the Atlantic Ocean, and the second in the Coral Triangle and Great Barrier Reef in the Indo-Pacific. Hierarchical clustering analysis of demosponges across the 12 marine realms revealed three groups. The Southern Ocean formed a single cluster, Eastern Tropical Pacific was paired with Temperate South America and the remaining realms merged into one group. This study, amplifies our understanding of global demosponge distribution, serving as a baseline to inform accordingly future conservation and management strategies.

keywords: species distribution modelling, biodiversity patterns, occurrence records, species richness, sponges, endemicity

Aknowledgements

The completion of this master thesis could not have been possible without the contribution and guidance of my supervisor, Professor Jorge Assis, and my colleague, Eliza Fragkopoulou. I am sincerely grateful for their support and invaluable advice in this scientific journey.

Table of contents

List of abbreviations

ADABoost: Adaptive Boosting

AUC: Area Under the Curve

BBNJ: Biodiversity Beyond National Jurisdiction

BioTIME: A database of biodiversity time series for the Anthropocene

BRT: Boosted Regression Trees

CBD: Convention of Biological Diversity

EMODnet: European Marine Observation and Data Network

ETP: Eastern Tropic Pacific

FAIR: principle of Findability, Accessibility, Interoperability and Reusability

GBIF: The Global Biodiversity Information Facility

GEBCO: General Bathymetric Chart of the Oceans

ICES: International Council of Exploration of the Sea

IUCN: International Union for Conservation of Nature

NBN: National Biodiversity Network

NOAA: National Oceanic and Atmospheric Administration

OBIS: Ocean Biodiversity Information System

SDM: Species Distribution Modelling

SDMs: Specie Distribution Models

TSS: True Skill Statistic

VIF: Variance Inflation Factor

WoRMS: World Register of Marine Species

XGBoost: Extreme Gradient Boosting

Chapter 1

General Introduction

1.1 Marine Biodiversity in the Anthropocene: Global Efforts and Conservation Priorities

In the Anthropocene, the intensification of human impact has led to unparalleled changes in the elements of the biosphere (de Repentigny, 2021). The global ocean is no exception, since human exploitation in the sea is accelerated, reaching deeper waters, and increasingly impacting marine life (Williams et al., 2010). Current major anthropogenic threats of marine diversity are overfishing, habitat loss and pollution (Costello et al., 2010). Furthermore, climate change, mediated by human activities, is introducing additional stressors such as sea level rise, ocean warming and acidification, nutrient load changes and altered ocean circulation (Brierley & Kingsford, 2009). These changes can result in cascading biodiversity losses, which in turn influence the functioning of natural ecosystems and jeopardizes human well-being (Johnson et al., 2017).

To address the biodiversity crisis, numerous global efforts have attempted to unite actions to stop or reverse the decline in biodiversity (Johnson et al., 2017). In 2002, countries signed the 2010 Biodiversity Target through the Convention of Biological Diversity (CBD) and later in 2010 adopted once again the Strategic Plan for Biodiversity 2011-2020 and Aichi targets to halt or reduce the worsening biodiversity trends (Xu et al., 2020). When these agreements failed to meet their targets by the designated time frame, a new framework was released from the Secretariat of the CBD in 2020, called post-2020 Global Biodiversity Framework (Xu et al., 2020). Several studies have tried to identify the reasons behind this failure and provide advice and recommendations for the implementation to meet the new global targets (e.g., Xu et al., 2020; Perino et al., 2022).

Anthropogenic impacts, compounded by the imperative for conservation planning and management following the post-2020 framework have further promoted the study of diversity trends and processes from regional to local levels (Tittensor et al., 2010). These patterns of biodiversity differ across biomes, taxa and scales, influenced by ecological and evolutionary processes over geological times (Gagné et al., 2020; Perino et al., 2022). In the marine environment, however, our understanding of these global patterns is limited. The most significant worldwide collaborative effort in this regard has been the Census of marine life, which aimed to assess and elucidate these patterns (Costello et al., 2010; Tittensor et al., 2010). Nevertheless, establishing appropriate baselines is crucial to follow future biodiversity shifts, since without them, we cannot accurately determine what and how much has been lost or evaluate the effectiveness of conservation and restoration efforts (Lin et al., 2022).

Understanding species distributions is fundamental to almost every aspect of managing biodiversity and two major fundamental metrics arise: species richness and endemism (Franklin, 2010; Selig et al., 2014). Species Distribution Models (SDMs) are well established statistical tools for mapping marine species ranges, unveiling endemicity and richness trends, from regional to global scales, an essential step to prioritize areas for conservation (Costello & Chaudhary, 2017). Designated priority areas are shaped by both spatial patterns and human impacts. In particular, regions marked by high endemicity, characterized by species with small geographic ranges, or areas of concentrated high richness, affecting many species, are given precedence in conservation initiatives (Tittensor et al., 2010; Costello & Chaudhary, 2017).

1.2 Species Distribution Modelling

The prediction of species distribution is vital for diverse applications in ecology, evolution, and conservation science (Elith et al., 2006). The foundations that conceptualized Species Distribution Modelling (SDM) date back in the last century, defining a species' niche within both geographic and environmental contexts (Colwell and Rangel, 2009; Sofaer et al., 2019). Today, SDMs have evolved into a fundamental tool across terrestrial, freshwater, and marine environments (Elith & Leathwick, 2009), along with the increasing availability of statistical methods, digital biological and environmental data that can be built in a geographic system (Miller, 2010).

SDMs are statistical computational tools that quantify the relationship between a species distribution and a set of environmental predictor variables (Miller, 2010; Franklin, 2010). By associating observations of species occurrence or abundance with environmental estimates, they deploy various statistical learning methods, providing insights into species distributions (Elith & Leathwick, 2009; Franklin, 2013). These models serve as pivot tools in research, offering a multitude of applications that are based on species distributions. They facilitate the investigation of related biogeographical variables, such as species richness and assemblages by stacking individual SDMs (Dubuis et al., 2011; Calabrese et al., 2014; Zurell et al., 2020).

Further, they help in evaluating invasiveness and proliferation potential (Srivastava et al., 2019) and understanding disease transmission (Peterson, 2006), detecting rare species sites (Jeliazkov et al., 2022), identifying potential translocation sites for endangered species (Eyre et al., 2022), testing evolution hypotheses (Graham et al., 2004) and guiding conservation and management decisions (Sofaer et al., 2019).

Given the diverse applications of SDMs, there has been a surge in literature focused on developing metrics and methods to improve species distributions predictions. The basic elements of the SDMs, georeferenced data, environmental predictors and relevant algorithms are all meticulously selected and curated accordingly. The utility of the models depends on the assessment of their performance, which pinpoints the strengths and limitations, thereby determining the range of their practical applications (Liu et al., 2009). However, several methodological challenges remain unresolved, leading to the development of various methods and approaches (Segurado et al., 2006).

The nature of biological data, ordinal (e.g., presence/absence), nominal (e.g., ranked abundance), and ratio (e.g., abundance, richness) levels, dictates the choice of algorithms, influencing the output of the SDM product (e.g., probability of occurrence, expected mean) (Miller, 2010). SDMs can be divided into two groups, those who require only presence data and those necessitating both presence and absence data, with the second group performing better (Brotons et al., 2004). In any case, initial presence data need to be unbiased both geographically and climatically (Barbet-Massin et al., 2012). Since true absence data are difficult to obtain and verify, pseudo-absences or background data serve as substitutes (Barbet-Massin et al., 2012). Another challenge is the adequate analytical power and thus the determination of a minimum number of records to generate an accurate model (van Proosdij et al., 2015). In such spatial analyses, the simple count might not be a good estimator of effective sample size, since the same number of close-distance observations convey less than widespread ones (Segurado et al., 2006). This phenomenon, called spatial autocorrelation, leads to overestimation of the effective sample size, and a variety of methods have been suggested to counteract its effects (Dormann et al., 2007). Additional concerns encompass the ratio of presences/absences (prevalence), the optimal number of pseudo-absences replicate sets, and the methods generate pseudo-absences (Barbet-Massin et al., 2012).

Environmental data, the second vital component of SDMs, have become increasingly available, with access to high-resolution layers derived from remotely sensed data, data interpolation or modelled data (Melo-Merino et al., 2020; Elith et al., 2006). In marine environments, the depth creates a third vertical dimension that exerts an additional challenge in modelling, especially for pelagic and high mobile species (Dambach & Rödder, 2011; Melo-Merino et al., 2020). The relevance and number of the environmental predictors, capturing a species' niche are essential in model designing, and a mix of biological and statistical criteria help determine the best set of these variables (Melo-Merino et al., 2020). Additionally, the environmental predictors available in GIS layers are structured in cells as grids with various resolution levels (Huby et al., 2007; Miller et al., 2010). The spatial resolution affects both the extent of the predicted suitable area and the relative importance of the environmental variables (Basher et al., 2014). Hence, layers need to be adjusted to ensure consistency in spatial extent, grid cell size and alignment, and projection (Elith et al., 2006). A multitude of environmental predictors is used to develop these models including among others, physical and chemical properties of the seawater, seabed topography, geographic variables (e.g., distance to a factor), anthropogenic pressures, and even distribution maps of other species to represent potential biotic interactions (e.g., predation) (Miller, 2010; Melo-Merino et al., 2020). However, their temporal resolution is often overlooked, even though is crucial for mobile species, as their distribution is driven by dynamic oceanographic processes that vary across time scales (Mannocci et al., 2017).

Selecting an appropriate machine learning algorithm is a pivotal consideration in modelling. Numerous SDMs exist to forecast species distributions by analyzing the relationship between occurrence and environmental predictors. Nevertheless, choosing the right algorithm tailored to specific data and objectives is challenging for researchers (Li & Wang, 2013). Despite the plethora of available methods and implementations, existing to fit these mathematical functions, no definitive consensus has emerged regarding which algorithms are optimal in specific contexts or across all scenarios (Araújo et al., 2019; Sofaer et al., 2019). To improve performance of SDMs, recent studies favor a multi-model ensemble strategy, which offers a framework to include relevant model features such as degree of uncertainty and agreement levels between results (Jones & Cheung, 2015; Robinson et al., 2017; Sofaer et al., 2019).

SDMs are underpinned by three fundamental goals: realism, generality, accuracy (Li & Wang, 2013; Araújo et al., 2019). Realism pertains to the model's ability to pinpoint key predictors impacting the system and to accordingly depict their effects and interactions. Precision is the model's ability to correctly forecast within the system being modeled (space and time). Generality is the model's ability to make predictions beyond its initial system, via projection or transfer to a different resolution and across time and space (transferability). (Araújo et al., 2019). Accomplishing all these three goals at the same time is challenging, many times requiring a sacrifice on one objective to fulfill the other two, depending on the study (Araújo & Peterson, 2012; Li & Wang, 2013). For instance, in climate change scenario predictions, generality is essential, and realism is important (Araújo et al., 2019). Hence, quantitative validation measures are necessary for assessing model performance and/or accuracy, precision, specificity, and sensitivity. Yet only a limited number of studies involve metrics of error or uncertainty, even though models accompanied by performance estimates foster trust in the predictions offered. (Robinson et al., 2017).

1.3 Sponges

Sponges, an ancient metazoan phylum, are an integral part of sessile benthic communities, having colonized all aquatic ecosystems, from freshwater to marine environments, intertidal to abyssal zones, tropics to polar zones (Barnes, 1999; Bell, 2008; Amina & Musayeib, 2018). They perform a multitude of functional roles regarding substrate, bentho-pelagic coupling, and associations with other organisms (Bell, 2008). Sponges interact with the substrate by either eroding or stabilizing it (Bell, 2008). Additionally, their feeding strategy as major filter-feeders, enables them to process large amounts of water, removing particulate and dissolved organic carbon and nutrients from the water column and thus rending carbon flow to higher trophic levels through predation (Bell, 2008; Bell et al., 2018).

Sponges form relationships with other organisms that can be either microscopic or macroscopic. They establish close partnerships with a variety of microorganisms, which are collectively called sponge holobiont, that contribute to their nutrition, defense, immunity, and host development, thus jointly impacting their health and functioning (Moitinho-Silva et al., 2017; Pita et al., 2018). The holobiont's nutritional strategies can be phototrophic, deriving energy primarily from photosynthesis, and thus contributing to primary production, or heterotrophic depending primarily on suspension feeding (Bell, 2008; Bell et al., 2018). This distinction is important as it could shape how climate change impacts sponges (Bennett et al., 2017). Furthermore, sponges with their associated microorganisms, produce secondary metabolites as defensive mechanisms, rendering them as invaluable sources of natural products that have diverse uses in the medical and biotechnology sectors and pharmaceuticals (Mehbub et al., 2014). Sponges also support diverse macrofaunal communities, as with their three-dimensional special complexity increase the number of available microhabitats, provide shelter from predators, and serve as spawning and nursery grounds. This structural contribution underscores their role as ecosystem engineers, delivering a multitude of economic and ecological services. (Bell, 2008; Kennington et al., 2013; Idan et al., 2018).

Recent advancements have led to global estimates of marine diversity and endemicity across taxa (Tittensor et al., 2010; Selig et al., 2014; Costello et al., 2017; Costello & Chaudhary, 2017; Gagné et al., 2020; Melo-Merino et al., 2020). Most distributional research is focused on fish, mammals, mollusks, owing to economic and ecological significance of the first two groups and the conservation concerns associated with the third (Melo-Merino et al., 2020). However, sponges, despite their diverse roles and widespread presence in the aquatic ecosystems, are often neglected (Renard et al., 2018). Only a limited number is published regarding their sponge distribution using modelling practices (Melo-Merino et al., 2020). Existing studies often narrow their focus on specific species or geographic extents (e.g, Downey et al., 2012; Howell et al., 2016) or in association with other taxa such as corals (e.g, Rooper et al., 2014).

Sponges are susceptible to climate change and direct human activities such as trawl fishing, oil and gas activities (Culwick et al., 2020). Under the global concern for marine biodiversity loss, there is growing demand to determine current species distribution patterns to effectively identify and measure changes in biodiversity (Fromont et al., 2006). To address this need, this study attempted to estimate the global biodiversity patterns of marine demosponges, members of the Class Demospongiae, which comprises ~81 % of the total phylum (Morrow & Cárdenas, 2015). As a first step, species records from this class were gathered from diverse online repositories and were then cross-validated for taxonomic accuracy, as well as bathymetric and geographic distribution against established expert knowledge. This process yielded records for 649 species, that were then utilized to produce a modelling ensemble of three machine learning algorithms paired with relevant environmental drivers. Finally, species richness patterns were revealed, after stacking individual species maps, along with the relative contributions and thresholds of the predictors.

For the purposes of this master thesis, the research methodology is divided into two chapters. The first chapter is related to the collection and validation of the occurrence records, while the second focuses on the development of the Species Distribution Models and elucidating global patterns of demosponge distribution.

1.4 References

Amina, M., & Al Musayeib, N. M. (2018). Biological and medicinal importance of sponge. Biological Resources of Water.

Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. Ecology, 93(7), 1527-1539.

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., ... & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. Science advances, 5(1), eaat4858.

Barnes, D. K. (1999). High diversity of tropical intertidal zone sponges in temperature, salinity and current extremes. African Journal of Ecology, 37(4), 424-434.

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(2), 327-338.

Basher, Z., Bowden, D. A., & Costello, M. J. (2014). Diversity and distribution of deep-sea shrimps in the Ross Sea region of Antarctica. PLoS One, 9(7), e103195.

Bell, J. J. (2008). The functional roles of marine sponges. Estuarine, coastal and shelf science, 79(3), 341-353.

Bell, J. J., Bennett, H. M., Rovellini, A., & Webster, N. S. (2018). Sponges to be winners under near-future climate scenarios. Bioscience, 68(12), 955-968.

Bennett, H. M., Altenrath, C., Woods, L., Davy, S. K., Webster, N. S., & Bell, J. J. (2017). Interactive effects of temperature and pCO 2 on sponges: from the cradle to the grave. Global Change Biology, 23(5), 2031-2046.

Brierley, A. S., & Kingsford, M. J. (2009). Impacts of climate change on marine organisms and ecosystems. Current biology, 19(14), R602-R614.

Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 27(4), 437-448.

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. Global Ecology and Biogeography, 23(1), 99-112.

Colwell, R. K., & Rangel, T. F. (2009). Hutchinson's duality: the once and future niche. Proceedings of the National Academy of Sciences, 106(supplement_2), 19651-19658.

Costello, M. J., & Chaudhary, C. (2017). Marine biodiversity, biogeography, deep-sea gradients, and conservation. Current Biology, 27(11), R511-R527.

Costello, M. J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H., & Miloslavich, P. (2010). A census of marine biodiversity knowledge, resources, and future challenges. PloS one, 5(8), e12110.

Costello, M. J., Tsai, P., Wong, P. S., Cheung, A. K. L., Basher, Z., & Chaudhary, C. (2017). Marine biogeographic realms and species endemicity. Nature communications, 8(1), 1057.

Culwick, T., Phillips, J., Goodwin, C., Rayfield, E. J., & Hendry, K. R. (2020). Sponge density and distribution constrained by fluid forcing in the deep sea. Frontiers in Marine Science, 7, 395.

Dambach, J., & Rödder, D. (2011). Applications and future challenges in marine species distribution modeling. Aquatic Conservation: Marine and Freshwater Ecosystems, 21(1), 92-100.

de Repentigny, P. C. (2021). To the Anthropocene and beyond: the responsibility of law in decimating and protecting marine life. In Transnational Environmental Law in the Anthropocene (pp. 180-196). Routledge.

Downey, R. V., Griffiths, H. J., Linse, K., & Janussen, D. (2012). Diversity and distribution patterns in high southern latitude sponges. PLoS One, 7(7), e41672.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. P., & Guisan, A. (2011). Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. Diversity and Distributions, 17(6), 1122-1131.

Elith*, J., H. Gaham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., ... & E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29(2), 129-151.

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.

Eyre, A. C., Briscoe, N. J., Harley, D. K., Lumsden, L. F., McComb, L. B., & Lentini, P. E. (2022). Using species distribution models and decision tools to direct surveys and identify potential translocation sites for a critically endangered species. Diversity and Distributions, 28(4), 700-711.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30(5), 609-628.

Franklin, J. (2010). Mapping species distributions: spatial inference and prediction. Cambridge University Press.

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. Diversity and distributions, 19(10), 1217-1223.

Fromont, J., Vanderklift, M. A., & Kendrick, G. A. (2006). Marine sponges of the Dampier Archipelago, Western Australia: patterns of species distributions, abundance and diversity. Marine, Freshwater, and Wetlands Biodiversity Conservation, 363-382.

Gagné, T. O., Reygondeau, G., Jenkins, C. N., Sexton, J. O., Bograd, S. J., & Hazen, E. L. (2020). Towards a global understanding of the drivers of marine and terrestrial biodiversity. PLoS ONE, 15(2), e0228065.

Graham, C. H., Ron, S. R., Santos, J. C., Schneider, C. J., & Moritz, C. (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution, 58(8), 1781-1793.

Howell, K. L., Piechaud, N., Downie, A. L., & Kenny, A. (2016). The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management. Deep Sea Research Part I: Oceanographic Research Papers, 115, 309-320.

Huby, M., Owen, A., & Cinderby, S. (2007). Reconciling socio-economic and environmental data in a GIS context: An example from rural England. Applied Geography, 27(1), 1-13

Idan, T., Shefer, S., Feldstein, T., Yahel, R., Huchon, D., & Ilan, M. (2018). Shedding light on an East-Mediterranean mesophotic sponge ground community and the regional sponge fauna. Mediterranean Marine Science, 19(1), 84-106.

Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., ... & Henle, K. (2022). Sampling and modelling rare species: Conceptual guidelines for the neglected majority. Global change biology, 28(12), 3754-3777.

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. Science, 356(6335), 270-275.

Jones, M. C., & Cheung, W. W. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. ICES Journal of Marine Science, 72(3), 741-752.

Kenchington, E., Power, D., & Koen-Alonso, M. (2013). Associations of demersal fish with sponge grounds on the continental slopes of the northwest Atlantic. Marine Ecology Progress Series, 477, 217-230.

Li, X., & Wang, Y. (2013). Applying various algorithms for species distribution modelling. Integrative zoology, 8(2), 124-135.

Lin, H., Caley, M. J., & Sisson, S. A. (2022). Estimating global species richness using symbolic data meta-analysis. Ecography, 2022(3), e05617.

Liu, C., White, M., & Newell, G. (2009, July). Measuring the accuracy of species distribution models: a review. In Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia (Vol. 4241, p. 4247).

Mannocci, L., Boustany, A. M., Roberts, J. J., Palacios, D. M., Dunn, D. C., Halpin, P. N., ... & Winship, A. J. (2017). Temporal resolutions in species distribution models of highly mobile marine animals: Recommendations for ecologists and managers. Diversity and Distributions, 23(10), 1098-1109.

Mehbub, M. F., Lei, J., Franco, C., & Zhang, W. (2014). Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. Marine drugs, 12(8), 4539-4577

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. Ecological Modelling, 415, 108837.

Miller, J. (2010). Species distribution modeling. Geography Compass, 4(6), 490-509.

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G. L., Cerrano, C., ... & Thomas, T. (2017). The sponge microbiome project. Gigascience, 6(10), gix077.

Morrow, C., & Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). Frontiers in Zoology, 12, 1-27.

Perino, A., Pereira, H. M., Felipe-Lucia, M., Kim, H., Kühl, H. S., Marselle, M. R., ... & Bonn, A. (2022). Biodiversity post-2020: Closing the gap between global targets and national-level implementation. Conservation letters, 15(2), e12848.

Peterson, A. T. (2006). Ecologic niche modeling and spatial patterns of disease transmission. Emerging infectious diseases, 12(12), 1822.

Pita, L., Rix, L., Slaby, B. M., Franke, A., & Hentschel, U. (2018). The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome, 6(1), 1-18.

Renard, E., Leys, S. P., Wörheide, G., & Borchiellini, C. (2018). Understanding animal evolution: the added value of sponge transcriptomics and genomics: the disconnect between gene content and body plan evolution. BioEssays, 40(9), 1700237.

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. Frontiers in Marine Science, 4, 421.

Rooper, C. N., Zimmermann, M., Prescott, M. M., & Hermann, A. J. (2014). Predictive models of coral and sponge distribution, abundance and diversity in bottom trawl surveys of the Aleutian Islands, Alaska. Marine Ecology Progress Series, 503, 157-176.

Segurado, P. A. G. E., Araujo, M. B., & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology, 43(3), 433-444.

Selig, E. R., Turner, W. R., Troëng, S., Wallace, B. P., Halpern, B. S., Kaschner, K., ... & Mittermeier, R. A. (2014). Global priorities for marine biodiversity conservation. PloS one, 9(1), e82898.

Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., ... & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. BioScience, 69(7), 544-557.

Srivastava, V., Lafond, V., & Griess, V. C. (2019). Species distribution models (SDM): applications, benefits and challenges in invasive species management.

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. Nature, 466(7310), 1098-1101.

van Proosdij, A. S., Sosef, M. S., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. Ecography, 39(6), 542-552.

Williams, M. J., Ausubel, J., Poiner, I., Garcia, S. M., Baker, D. J., Clark, M. R., ... & Halpin, P. N. (2010). Making marine life count: a new baseline for policy. PLoS biology, 8(10), e1000531.

Xu, H., Cao, Y., Yu, D., Cao, M., He, Y., Gill, M., & Pereira, H. M. (2021). Ensuring effective implementation of the post-2020 global biodiversity targets. Nature Ecology & Evolution, 5(4), 411-418.

Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. Journal of Biogeography, 47(1), 101-113.

Chapter 2

A global dataset of demosponge distribution records

Currently under review in "Data in Brief" scientific journal

2.1 Article information

Article title

A global dataset of demosponge distribution records

Authors

Ariadni Vafeiadou[1], Eliza Fragkopoulou[1] & Jorge Assis[1,2]

Affiliations

[1] Center of Marine Sciences (CCMAR-CIMAR), University of the Algarve, 8005-139 Faro, Portugal.

[2] Faculty of Bioscience and Aquaculture, Nord Universitet, Postboks 1490, Bodø, Norway.

Corresponding author's email address and Twitter handle

Jorge Assis: jorgemfa@gmail.com @jorgemfassis

Ariadni Vafeiadou: ariadne.vaf@gmail.com

Keywords

keywords: marine biodiversity, biodiversity information, biogeographic information, occurrence records, sponges

2.2 Abstract

Biodiversity information in the form of species occurrence records is key for monitoring and predicting current and future biodiversity patterns, as well as for guiding conservation and management strategies. However, the reliability and accuracy of this information are frequently undermined by taxonomic and spatial errors. Additionally, biodiversity information facilities often share data in diverse incompatible formats, precluding seamless integration and interoperability. We provide a comprehensive quality-controlled dataset of occurrence records of the Class Demospongiae, which comprises 81% of the entire Porifera phylum. Demosponges are ecologically significant as they structure rich habitats and play a key role in nutrient cycling within marine benthic communities. The dataset aggregates occurrence records from multiple sources, employs dereplication and taxonomic curation techniques, and is flagged for potentially incorrect records based on expert knowledge regarding each species' bathymetric and geographic distributions. It yields 376,242 records of 1,808 accepted demosponge species, that are provided under the FAIR principle of Findability, Accessibility, Interoperability and Reusability in the Darwin Core Standard. This dataset constitutes the most up-to-date baseline for studying demosponge diversity at the global scale, enabling researchers to examine biodiversity patterns (e.g., species richness and endemicity), and project potential distributional shifts under future scenarios of climate change.

2.3 Specifications table

Table 2.3.1. Specifications table regarding the research field these data address, type of data, and their sources.

Subject	Biodiversity
Specific subject area	Marine macroecology, marine biogeography, biodiversity data, marine conservation and management, climate change assessments
Type of data	Table Chart Graph Figure

How the data were acquired	Occurrence records and expert knowledge were acquired from biodiversity information facilities. Data were processed using R statistical computing software, version 4.2.2 (2023)
Data format	Excel files (Raw and Filtered)
Description of data collection	Georeferenced occurrence records of the class Demospongiae, dereplicated, taxonomically curated, flagged for potentially incorrect entries regarding each species' bathymetric and geographic distributions based on expert knowledge available in major databases of biological traits, and standardized with Darwin Core Standard.
Data source location	Institution: CCMAR- Centre of Marine Sciences City/Town/Region: Faro, Algarve Country: Portugal Occurrence records of demosponge species compiled from the biodiversity information facilities: (1) Ocean Biodiversity Information System (https://obis.org) (2) Global Biodiversity Information Facility (https://www.gbif.org) (3) Deep-Sea Coral & Sponge Map Portal, National Oceanic and Atmospheric Administration (https://www.ncei.noaa.gov/maps/deep-sea-corals/mapSites.htm) (4) National Biodiversity Network, NBN atlas (https://nbnatlas.org/) (5) Vulnerable Marine Ecosystems, International Council for the Exploration of the Sea (https://vme.ices.dk/download.aspx) (6) PANGAEA – Data Publisher for Earth & Environmental Science (https://www.pangaea.de) (7) BioTIME, A database of biodiversity time series for the Anthropocene (https://biotime.st-andrews.ac.uk) (8) Integrated Digitized Biocollections (https://www.idigbio.org) (9) European Marine Observation and Data Network (EMODnet) – Data Ingestion Portal (https://www.emodnet-ingestion.eu/) (10) Aquamaps, a global online database containing standardized distribution maps for marine species (https://www.aguamaps.org) Expert knowledge of demosponge species compiled from the biodiversity information facility: (10) Aquamaps, a global online database containing standardized distribution maps for marine species (https://www.aguamaps.org) (11) SeaLifeBase, a global online database of information about marine life (https://www.sealifebase.ca)

2.4 Value of the data

- The most up-to-date dataset of demosponge distribution records at a global scale. Marine sponges are keystone components of marine benthic communities, by serving as habitat builders, promoting biodiversity, and influencing nutrient cycling (Bell, 2008). Additionally, they constitute a valuable source of natural products with various applications in biomedicine, pharmaceuticals, and biotechnology (Mehbub, et al., 2014). Yet, sponges face numerous threats from environmental changes and human activities, including deep-sea industrialization and fishing.

- The dataset is curated, with dereplicated and taxonomically standardized records, flagged for potentially incorrect records, and provided under the FAIR principle in Darwin Core Standard, for seamless integration in statistical analyses and interoperability across biodiversity datasets.

- The dataset is a baseline to describe species' distributions at the global scale and address niche-based questions, which comprise projections of climate-induced range shifts across space and time.

- The dataset can assist researchers in tackling priority questions associated with demosponges macroecology, biogeography and climate change responses and impacts, which together can support the implementation of well-informed strategies for conserving, managing and restoring marine biodiversity.

2.5 Objective

Macroecology, biogeography and conservation research require complete and accurate occurrence data describing the distribution of species (Aubry et al., 2017). Although open-access biodiversity databases like the Ocean Biodiversity Information System (OBIS, 2023) provide access to such information, they often contain spatial and taxonomic errors and can be incomplete. Additionally, the presence of duplicated data in various formats hampers seamless integration and interoperability (Assis et al., 2020). Here, we provide a dataset of demosponge distribution records at the global scale, comprising dereplicated records of 1,808 taxonomically standardized species and incorporating a quality control system flagging potentially incorrect records (Vafeiadou et al., 2023; in prep.). Data are made available under the FAIR principle of Findability, Accessibility, Interoperability and Reusability in the Darwin Core Standard (Darwin Core Maintenance Group, 2021)

2.6 Data description

The dataset of occurrence records of species belonging to the class Demospongiae is provided in Excel format. Rows refer to occurrence records and columns are compatible with the data fields of Darwin Core Standard (Darwin Core Maintenance Group, 2021), with a focus on the date, source, location of records, taxonomy, and finally quality flag of records (Table 2.6.1).

Table 2.6.1. Data fields of the global dataset of demosponge distribution records (Additional information on Darwin Core Standard (Darwin Core Maintenance Group, 2021): https://dwc.tdwg.org).

Field	Description
aphiaID	Identifier of the taxon, linked to the World Register of Marine Species
name	Name of the taxon, as originally reported
acceptedName	Accepted name of the taxon, retrieved from the World Register of Marine Species
kingdom	Higher taxonomic classification
phylum	Higher taxonomic classification
class	Higher taxonomic classification
order	Higher taxonomic classification
family	Higher taxonomic classification
genus	Higher taxonomic classification
decimalLongitude	Geographical longitude in decimal degrees of the record’s location
decimalLatitude	Geographical latitude in decimal degrees of the record’s location

coordinateUncertaintyInMeters	Distance (in meters) from the decimal Latitude and decimal Longitude that describes the center of the circle containing the record’s location
depthAccuracy	Depth uncertainty of the record (in meters), as originally reported
locality	Name of the record’s location
minimumDepthInMeters	Minimum depth of the record (in meters), as originally reported
maximumDepthInMeters	Maximum depth of the record (in meters), as originally reported
year	Four-digit year in which the observation occurred
month	Two-digit month in which the observation occurred
day	Two-digit day in which the observation occurred
bibliographicCitation	Bibliographic reference of the record
license	“A legal document giving official permission to do something with the resource”
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties.
scientificNameAuthorship	Authorship information for the scientificName
taxonomicStatus	The status of the use of the scientificName as a label for a taxon.
coordinatePrecision	A decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude.
country	The name of the country or major administrative unit in which the Location occured
individualCount	The number of individuals represented present at the time of the Occurrence.
basisOfRecord	The specific nature of the data record.
measurementOrFact	Quality control based on the flagging system: flagGeographicRange ‘−1’ for records outside the known geographic distribution of species flagVerticalRange ‘−1’ for records outside the known depth range of species flagLand ‘−1’ for records over land

Initially, 506,492 records of occurrence of species belonging to the class Demospongiae were gathered from online biodiversity databases. Records were taxonomically standardized using the World Register of Marine Species, and duplicated and non-georeferenced records were removed. Expert knowledge on the vertical and geographic distribution of species belonging to the class Demospongiae was gathered from the SeaLifeBase (Palomares & Pauly, 2023), an online database with information about marine life, and Aquamaps (Kaschner et al., 2019), a database providing expert-curated species range maps. Occurrence records falling outside the known vertical and geographical distribution, as well as on land, were then flagged as potentially incorrect, resulting in a final pruned dataset with 376,242 records of 1,808 species belonging to 270 genera, 87 families and 21 orders of the Class Demospongiae (Table 2.6.2, Figure 2.6.1) and covering the period from 1786 to 2023 (Figure 2.6.2) (WoRMS Editorial Board, 2023).

Table 2.6.2 Number of species, records and flagged records falling out of the known (1) bathymetric and (2) geographic distribution, and (3) over land. Numbers in parentheses represent percentages.

Order	Species number	Records	Flagged		
			Outside vertical range	Outside distribution range	On land
Agelasida	26	5308	3294 (62)	728 (14)	3 (0.06)
Axinellida	115	27250	12880 (47.27)	4112 (15.09)	109 (0.4)
Biemnida	30	1684	713 (42.34)	348 (20.67)	5 (0.3)
Bubarida	38	4356	2868 (65.84)	542 (12.44)	19 (0.44)
Chondrillida	17	5667	1421 (25.07)	1551 (27.37)	140 (2.47)
Chondrosiida	5	3035	47 (1.55)	523 (17.23)	52 (1.71)
Clionaida	72	45552	4073 (8.94)	6952 (15.26)	297 (0.65)
Dendroceratida	27	4269	789 (18.48)	1472 (34.48)	35 (0.82)
Desmacellida	9	97	31 (31.96)	38 (39.18)	0 (0)
Dictyoceratida	106	39943	16967 (42.48)	7210 (18.05)	206 (0.52)
Haplosclerida	300	36084	18359 (50.88)	6921 (19.18)	355 (0.98)
Merliida	5	64	12 (18.75)	15 (23.44)	0 (0)
Poecilosclerida	505	71955	19353 (26.9)	16167 (22.47)	323 (0.45)
Polymastiida	34	18597	13876 (74.61)	2766 (14.87)	48 (0.26)
Scopalinida	13	2513	792 (31.52)	524 (20.85)	3 (0.12)
Sphaerocladina	1	13	11 (84.62)	0 (0)	0 (0)
Suberitida	123	57089	13093 (22.93)	17604 (30.84)	949 (1.66)
Tethyida	46	6458	2309 (35.75)	1578 (24.43)	47 (0.73)
Tetractinellida	305	36240	12911 (35.63)	11622 (32.07)	76 (0.21)
Trachycladida	2	164	20 (12.2)	139 (84.76)	2 (1.22)
Verongiida	29	9904	3850 (38.87)	2589 (26.14)	152 (1.53)
Total	1808	376242	127669	83401	2821

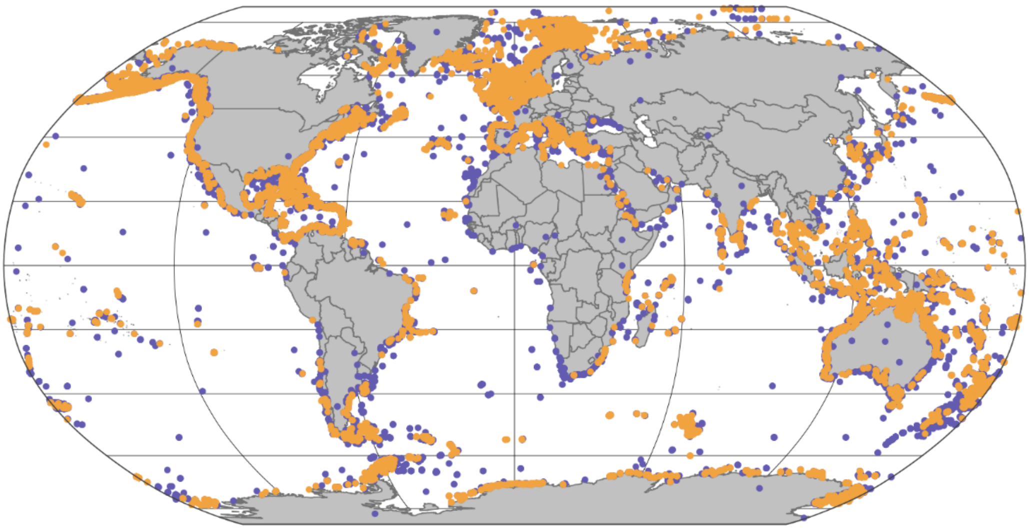


Figure 2.6.1. Global map of demosponge records. Points in purple indicate those flagged as potentially wrong, based on their known vertical and bathymetric ranges and/or on land. Points in orange demonstrate occurrences that are flagged as correct.

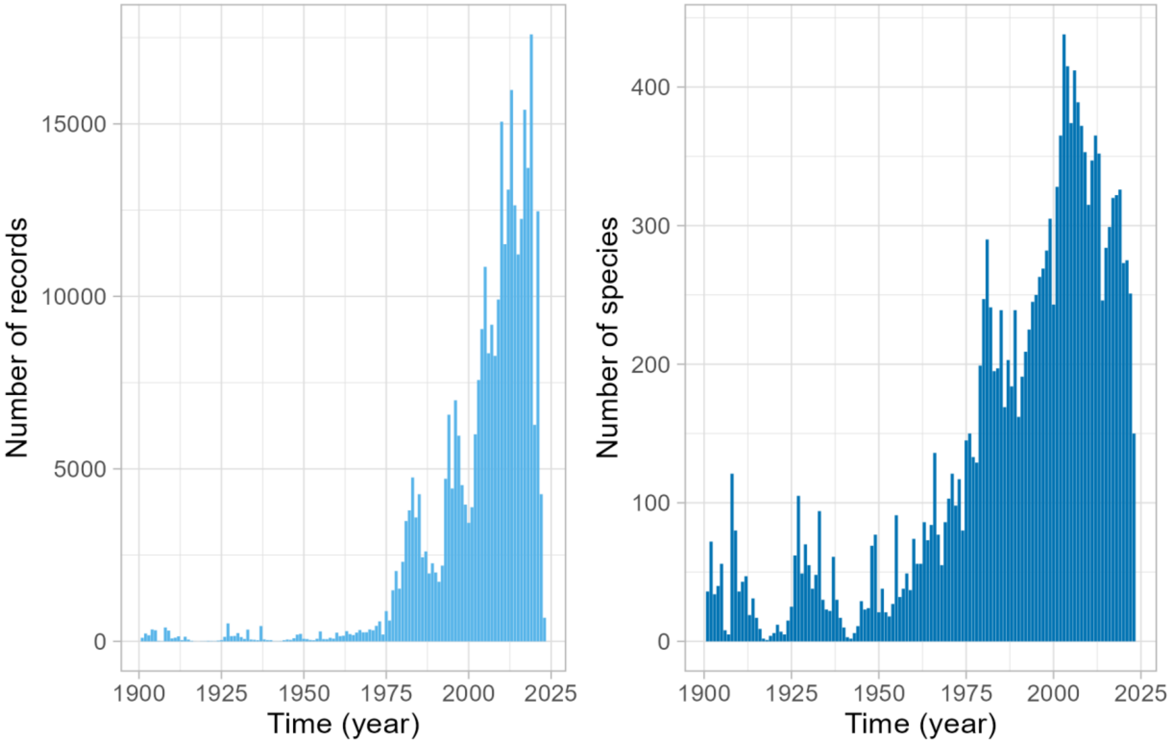


Figure 2.6.2. Number of demosponge (a) records and (b) species available in the demosponge dataset per year from 1900 to 2023 (data are available since the year 1786).

2.7 Experimental design, materials and methods

The collection and curation steps of the global dataset of demosponge distribution records are detailed below.

Step 1. Collating the list of sponge species belonging to the Class Demospongiae
The taxonomy of sponges covers a broad spectrum of species. The scope of this dataset is placed on marine species of the class Demospongiae, the largest sponge class comprising 81% of all sponges (Morrow & Cárdena, 2015). A list of taxonomically accepted species of the class Demospongiae was collated from the World Register of Marine Species (WoRMS) (WoRMS Editorial Board, 2023) and was used to search for occurrence records. WoRMS is an authoritative reference system for marine species that offers a unique identification code (aphiaID) associated with a standardized accepted name, and related taxonomic information.

Step 2. Occurrence records collection
Occurrence records of the targeted species were collected from major online biodiversity databases: (1) Ocean Biodiversity Information System (OBIS, 2023), (2) Global Biodiversity Information Facility (GBIF, 2023), (3) Deep-Sea Coral & Sponge Map Portal, National Oceanic and Atmospheric Administration (NOAA, 2023), (4) National Biodiversity Network, NBN atlas (NBN Trust, 2023), (5) Vulnerable Marine Ecosystems, International Council for the Exploration of the Sea (ICES, 2023), (6) PANGAEA – Data Publisher for Earth & Environmental Science (PANGAEA, 2023), (7) BioTIME, A database of biodiversity time series for the Anthropocene (BioTIME, 2023), (8) Integrated Digitized Biocollections (Integrated Digitized Biocollections, 2023), (9) European Marine Observation and Data Network, Data Ingestion Portal (EMODnet, 2023), (10) Aquamaps (Kaschner et al., 2019). The original source of each record is reported in the respective fields of the Darwin Core Standard.

Only records reporting no copyright and without restrictions for any use or any use with appropriate attribution (e.g., CC0 or CC BY, www.creativecommons.org) were stored in the dataset.

Step 3. Taxonomic curation

Taxonomic standardization was performed for each entry with the WoRMS (WoRMS Editorial Board, 2023). Entries with status other than accepted were matched with the currently valid species names. Records were also checked to belong to the Demospongiae class, and if not, they were discarded from the dataset.

Step 4. Pruning of occurrence records

Records lacking coordinated information were discarded from the dataset. Additionally, duplicated records sharing the same spatial (longitude, latitude, depth) and temporal information (year, month, day) were discarded from the dataset.

Step 5. Quality control flagging of occurrence records

In marine biodiversity research, quality control is an essential step, especially when handling thousands of records. No matter how reliable a source is, the presence of incorrect records is still possible and can be propagated across repositories due to automatic interoperability (Assis et al., 2020). To solve this problem, a quality control system was followed as outlined by Assis et al., 2020 to flag records over land and/or with potentially wrong geographical and vertical distributions.

Records over land were identified with a polygon provided by Natural Earth (Natural Earth, 2020), a public domain map that encompasses different scales. Here, the 1:10 m scale layer was employed as a reference. The criterion for flagging records was based on a 1km Euclidean distance from the ocean, as in Assis et al., 2020.

Additionally, the depth of each record was extracted based on the General Bathymetric Chart of the Oceans (GEBCO), a global terrain model providing elevation data, in meters, on a 15 arc-second interval grid (GEBCO Compilation Group, 2022). The depth values were compared to the known bathymetric distribution of the corresponding species based on expert knowledge information provided by SeaLifeBase (Palomares & Pauly, 2023) and Aquamaps (Kaschner et al., 2019). More specifically, records were flagged when their depth values fell out of their known bathymetric range. Likewise, the validation of geographical locations, based on longitude and latitude, was compared to the expert knowledge information for the corresponding species provided by SeaLifeBase (Palomares & Pauly, 2023) and Aquamaps

(Kaschner et al., 2019). Known geographical locations were reported in the form of Food and Agriculture Organization (FAO) Major Fishing Areas (FAO, 2023).

Step 6. Dataset format standardization

The dataset was standardized based on the Darwin Core Standard, a framework that provides identifiers, labels, and specific definitions to facilitate sharing of information about biodiversity (Darwin Core Maintenance Group, 2021). The dataset provides standardized information for each entry, on source, taxonomy, date, location, depth and quality flag (Table 2.6.1).

Potential use of the dataset

The global dataset of demosponge distribution records is an up-to-date baseline for characterizing the distribution of species and unravelling biodiversity patterns such as endemicity centres and species richness hotspots to guide conservation, management and restoration strategies (Fragkopoulou et al., 2022). It can also be used in modelling frameworks to identify suitable habitats of overlooked species (Boavida et al., 2016) and project future range shifts under climate change scenarios (Assis et al., 2014).

Ethics statements

The present work complies with ethical requirements and does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT author statement

Conceptualization: VA, EF, and JA. Data curation: VA and JA. Funding acquisition: JA and ES. VA and EF led the writing with support from JA. All authors contributed to editing the article and approved the submitted version.

Acknowledgments

This study received Portuguese national funds from the Horizon Europe Framework Programme through project MPAEurope (HORIZON-CL6-2021-BIODIV-01-12) and the Foundation for Science and Technology (FCT) through programs UIDB/04326/2020, UIDP/04326/2020, LA/P/0101/2020 and PTDC/BIA-CBI/6515/2020, the Individual Call to Scientific Employment Stimulus 2022.00861 to J.A. and the fellowship SFRH/BD/144878/2019 to E.F

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

2.8 References

Aubry, K. B., Raley, C. M., & McKelvey, K. S. (2017). The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS One*, *12*(6), e0179152.

Assis, J., Fragkopoulou, E., Frade, D., Neiva, J., Oliveira, A., Abecasis, D., ... & Serrão, E. A. (2020). A fine-tuned global distribution dataset of marine forests. *Scientific Data*, *7*(1), 119.

Assis, J., Serrao, 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*(11), 2797-2810.

Bell, J. J. (2008). The functional roles of marine sponges. *Estuarine, coastal and shelf science*, *79*(3), 341-353.

BioTIME. (2023). A database of biodiversity time series for the Anthropocene. https://biotime.st-andrews.ac.uk

Boavida, J., Assis, J., Silva, I., & Serrão, E. A. (2016). Overlooked habitat of a vulnerable gorgonian revealed in the Mediterranean and Eastern Atlantic by ecological niche modelling. *Scientific Reports*, *6*(1), 36460.

Darwin Core Maintenance Group. (2021). Darwin Core list of terms. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/list/2023-09-18

European Marine Observation and Data Network (EMODnet). (2023). Data Ingestion Portal. Retrieved from https://www.emodnet-ingestion.eu/

FAO. (2023). FAO Major Fishing Areas. Fisheries and Aquaculture Division [online]. Rome. https://www.fao.org/fishery/en/collection/area

Fragkopoulou, E., Serrão, E. A., De Clerck, O., Costello, M. J., Araújo, M. B., Duarte, C. M., ... & Assis, J. (2022). Global biodiversity patterns of marine forests of brown macroalgae. *Global Ecology and Biogeography*, *31*(4), 636-648.

GBIF: The Global Biodiversity Information Facility. (2023). What is GBIF? Available from https://www.gbif.org/what-is-gbif [13 January 2020].

GEBCO Compilation Group. (2022). GEBCO 2022 Grid. https://doi.org/10.5285/e0f0bb80-ab44-2739-e053-6c86abc0289c

Integrated Digitized Biocollections. (2023). Retrieved from https://www.idigbio.org

International Council for the Exploration of the Sea. (2023). Vulnerable Marine Ecosystems. https://vme.ices.dk/download.aspx

Kaschner, K., Kesner-Reyes, K., Garilao, C., Segschneider, J., Rius-Barile, J., Rees, T., & Froese, R. (2019). AquaMaps: Predicted range maps for aquatic species. https://www.aquamaps.org

Mehbub, M. F., Lei, J., Franco, C., & Zhang, W. (2014). Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. *Marine drugs*, *12*(8), 4539-4577.

Morrow, C., & Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). *Frontiers in Zoology*, *12*, 1-27.

National Oceanic and Atmospheric Administration. (2023). Deep-Sea Coral & Sponge Map Portal. https://www.ncei.noaa.gov/maps/deep-sea-corals/mapSites.htm

Natural Earth. (2023). Retrieved from https://www.naturalearthdata.com

NBN Trust. (2023). The National Biodiversity Network (NBN) Atlas. https://ror.org/00mcxye41.

Ocean Biodiversity Information System (OBIS). (2023). Retrieved from https://obis.org

Palomares, M.L.D., & Pauly, D. (2023). SeaLifeBase. World Wide Web electronic publication. http://www.sealifebase.org, version (04/2023).

PANGAEA – Data Publisher for Earth & Environmental Science. (2023). Retrieved from https://www.pangaea.de

Vafeiadou, A., Fragkopoulou, E., & Assis, J. (2023). Global demosponge diversity dataset [Dataset]. Figshare. link to be published.

WoRMS Editorial Board. (2023). World Register of Marine Species. VLIZ. https://www.marinespecies.org. doi:10.14284/17

Chapter 3

Global biogeographic patterns of marine sponges

3.1 Article information

Article title

Global biogeographic patterns of marine sponges

Authors

Ariadni Vafeiadou[1], Eliza Fragkopoulou[1] & Jorge Assis[1,2]

Affiliations

[1] Center of Marine Sciences (CCMAR-CIMAR), University of the Algarve, 8005-139 Faro, Portugal.

[2] Faculty of Bioscience and Aquaculture, Nord Universitet, Postboks 1490, Bodø, Norway.

Corresponding author's email address and Twitter handle

Jorge Assis: jorgemfa@gmail.com @jorgemfassis

Ariadni Vafeiadou: ariadne.vaf@gmail.com

Keywords

keywords: species distribution modelling, biodiversity patterns, occurrence records, species richness, sponges, endemicity

3.2 Abstract

Anthropogenic activities have dramatically impacted our oceans, which have triggered cascading biodiversity loss. This, in turn, affects ecosystem functioning, services, and ultimately human well-being. In response, international agreements set conservation goals for positive biodiversity. A prerequisite for effective management is a comprehensive understanding of current marine species distribution. Sponges are essential components of benthic communities, having colonized all aquatic ecosystems. Despite their diverse ecological roles, sponges are often overlooked. In this study, we focused on marine sponges from the class Demospongiae, which make up ~ 81% of all sponges. Using Species Distribution Models, we delineated global biodiversity patterns of marine demosponges. Occurrence records, taxonomically and spatially curated, were paired with relevant environmental predictors using an ensemble of machine learning algorithms. The resulting individual models were stacked to unravel biodiversity trends on a global scale, along with overall relative predictors' contribution and their corresponding critical thresholds. The findings showed that demosponge habitats are predominantly influenced by thermal conditions, pH, and phytoplankton productivity. Highest diversity was estimated in the tropical and subtropical zones, with one hotspot residing in the Caribbean Sea and Gulf of Mexico in the Atlantic Ocean, and the second in the Coral Triangle and Great Barrier Reef in the Indo-Pacific. Hierarchical clustering analysis of demosponges across the 12 marine realms revealed three groups. The Southern Ocean formed a single cluster, Eastern Tropical Pacific was paired with Temperate South America and the remaining realms merged into one group. This study, amplifies our understanding of global demosponge distribution, serving as a baseline to inform accordingly future conservation and management strategies.

3.3 Introduction

Sponges (Porifera) are ancient sessile metazoans with a fossil record dating back 541 million years, during which some member of this phylum flourished or waned, with regards to past climatic fluctuations that caused evolutionary opportunities (Muir et al., 2017). They are major components of benthic communities thriving in a wide range of marine habitats (Barnes, 1999). From the intertidal to abyssal zone, tropics to polar zones, across varying salinity, substratal and temperature ranges, sponges come in various colors, shapes and sizes (Barnes, 1999; Amina & Musayeib, 2018; Bell, 2018). Their importance is reflected in their diverse ecological roles. They act as habitat builders, forming three-dimensional structures, influencing stability, and increasing the available microhabitats to promote biodiversity, offer shelter from predators, and serve as areas of spawning and nursery (Idan et al., 2018). Being active filter-feeders, sponges are perceived as keystone species of bentho-pelagic coupling and nutrient cycling, since they remove carbon and nutrients from the water column and render them available to higher trophic levels (Bell, 2008; Pomponi et al.,2019). Additionally, they represent a valuable economic resource of natural products with diverse applications for biotechnology, medicine, and pharmaceuticals (Mehbub et al., 2014).

Despite the ecological importance of Porifera, this phylum is often overlooked, resulting in limited and fragmented knowledge regarding their biodiversity patterns at biogeographical scales. Existing studies have focused on specific habitats such as Caribbean sponges in coral reef communities (e.g. Diaz & Rützler, 2001), geographic areas such as Australian waters (e.g., Huang et al., 2011), certain depths (e.g., Pomponi et al., 2019), few individual species and/or certain assemblages (e.g, Carballo et al., 2008) or a mix of all those (Idan et al., 2018). Up to date, only one study has attempted to estimate the global diversity of this phylum but was based solely on sponge records, representing a minimal estimate of the distribution of already studied regions (van Soest et al., 2012). In addition, anthropogenic activities are also exerting influence on sponge survival, abundance, and distribution (Vad et al., 2021), thus shifting biodiversity baselines (Assis et al., 2013). Under this prism, and in a collective attempt to safeguard marine biodiversity, the International Union for Conservation of Nature (IUCN) Word Conservation Congress has called for the protection of at least 30% of the global ocean by 2030 (IUCN, 2016) and in June, 2023, the BBNJ Treaty (biodiversity beyond national jurisdiction) was adopted to establish large-scale marine protected areas in the high seas (Tiller & Mendenhall, 2023). This further highlights the need to estimate and identify biodiversity hotspots globally to inform conservation strategies and proper management.

In this study, we used machine learning Species Distribution Modelling (SDM) to help bridge the knowledge gap regarding sponge distribution on a global scale, aligning with the growing demand to inform ongoing conservation demands. Species Distribution Modelling is a widely used tool in ecology, biogeography, and conservation (Jiménez-Valverde et al., 2008), which is based on correlating species occurrence records with environmental predictor variables (Franklin, 2013; Elith et al., 2006), giving insights to ecologists and conservation managers of species' distribution patterns (Allouche et al., 2006). We provide an analysis of the global biodiversity patterns of 649 species of the class Demospongiae, which comprises ~ 81% of the sponge biome (Morrow & Cárdenas, 2015). To achieve this, we used a comprehensive dataset of occurrence records, which is dereplicated, and assessed both taxonomically and spatially by expert available knowledge (Vafeiadou et al., 2023; Chapter 2), and a set of meaningful predictors to fit three machine learning algorithms (Boosted Regression Trees, Adaptive Boosting, Extreme Gradient Boosting). We then stacked the individual predictive models to uncover global diversity trends among demosponges. Our research establishes foundational data, unveiling critical habitats of demosponges. This essential knowledge will help guide future conservation and management strategies, directing efforts towards areas important to demosponges, such those of high richness and endemicity. Identifying and protecting these areas, not only ensures the preservation of demosponges, but also secures the marine ecosystems they sustain.

3.4 Materials & Methods

Occurrence records and environmental predictors

Occurrence records of marine sponges that belong to the Class Demospongiae were collected from a curated dataset consisting of 1,808 accepted species (Vafeiadou et al., 2023; Chapter 2). This dataset contains records of demosponges collected from multiple online repositories, dereplicated and curated based on current taxonomic data available in World Register of Marine Species (WoRMS, 2023). The dataset cross-validates bathymetric and geographic distributions based on expert knowledge. Out of the 1,808 species, 649 were used in this study since SDM needs a sample size of at least five geographically distinct observations (van Proosdij et al., 2016).

Considering demosponges's biology and ecology, relevant environmental predictors related to physical, topographic and disturbance ocean properties, were selected to produce the models. The respective climate data were extracted from Bio-ORACLE v2.0 (Assis et al., 2017b), specifically, benthic layers of temperature (long-term minimum and maximum), dissolved molecular oxygen (long-term minimum), pH (long-term minimum), salinity (long-term minimum), seawater speed (long-term minimum), silicate (long-term minimum), total phytoplankton productivity (long-term minimum), slope (mean), terrain ruggedness index (mean).

Stacked Species Distribution Modelling

An ensemble of three high performance algorithms were utilized to produce individual species distribution models, namely, Boosted Regression Trees (BRT) (Elith & Leathwick, 2017), Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016) and Adaptive boosting (AdaBOOST) (Citores et al., 2020). A number of advantages comes with their use on species distribution modelling including their ability to excel in handling biodiversity datasets with low prevalence (Elith et al., 2006), effectively capture intricate relationships and interactions between predictor and response variables (Barbet-Massin et al., 2012) and reduce overfitting by means of hyperparameter tunning and the enforcement of monotonic responses (Hofner et al., 2011; Elith & Leathwick, 2017), a crucial step in enhancing model transferability (Elith et al., 2006; Elith & Graham, 2009; Heikkinen et al., 2012).

These machine learning tools require presence and absence data, but given the fact that only presences were available, pseudo-absences were randomly generated per species. A filtering procedure was applied to both occurrences and pseudo-absences to remove excess information and counteract the spatial autocorrelation and spatial bias in the models (Dormann et al., 2007). As a first step, one occurrence from the pool was randomly selected within a species-specific distance, where predictor variables did not demonstrate significant autocorrelation (Segurado et al., 2006; Di Cola et al., 2017). Built correlograms helped determining such distance by evaluating the Pearson's correlation coefficient of predictors as a function of geographic distance. The second step involved only the pseudo-absences, which were further pruned in a 1:1 ratio with occurrence records, a process followed by Barbet-Massin et al., 2012. To prevent redundancy in our modelling data, pseudo-absences were categorized based on climate conditions using K-means clustering algorithm with predictor variables, with the k parameter set equal to the number of occurrences (Senay et al., 2013). This extra step was essential in dealing with class imbalance, a significant concern in machine learning (Japkowicz & Stephen, 2002), which enhances our ability to isolate the relative contribution of predictor variables on model outcomes (Senay et al., 2013).

A cross-validation framework with ten random folds was followed to estimate predictive error and tune model parameters, with the use of independent spatial blocks represented by hexagons, with their dimension aligned with the previously inferred climatic uncorrelated distances (Elith et al., 2008; Roberts et al., 2017; Valavi et al., 2019). This approach involves testing all the hyperparameter combinations for each algorithm, a process referred to as the "grid-search" method. Regarding BRT, the learning rate was set from 0.1 to 0.01, at a step of 0.001, tree complexity from 1 to 4, and number of trees from 50 to 100, at a step of 50 (Assis et al., 2017a). Accordingly, for XGBoost, gamma was set from 0 to 5, interaction depth from 1 to 4, shrinkage from 0.1 to 0.5 at a step of 0.1 and number of rounds from 10 to 100 at a step of 10 (Valavi et al., 2022). For ADABoost, the number of interactions ranged from 50 to 250 (step 50), degrees of freedom varied from 1 to 12 and shrinkage from 0.25 to 1 (step 0.25) (Krause-Jensen et al., 2020). For the training of the competitive models, data were divided into nine random folds, reserving one-fold at a time for testing to evaluate predictive error. To address overfitting and improve model generalization, monotonicity constraints were applied to ensure a positive or negative fit for each predictor based on anticipated biological impact on the model's response (Hofner et al., 2011). All variables were considered to have a positive effect on the response of the models, apart from maximum temperature. Between predictors, Pearson's correlation coefficient (r) and Variance Inflation Factor (VIF) were determined.

Three metrics were used to evaluate predictor performance; Boyce index which is particularly suitable for presence-only models, and ranges between -1 to 1, true skill statistic (TSS) which ranges between 0 and 1 and area under the curve (AUC) that ranges between 0 and 1 (Boyce et al., 2002; Allouche et al., 2006; Bellard et al., 2013). Values closer to 1 indicate predictions matching observed patterns and values closer to 0 or less indicate a performance no better than random, while specifically for AUC a score higher than 0.8 indicates good performance (Fielding & Bell, 1997; Allouche et al., 2006; Bellard et al., 2013).

Models were created, encompassing all predictors, and employing hyperparameter combinations that achieved lower prediction error in cross-validation, to investigate the relative contribution of predictors and to generate partial dependency plots illustrating each predictor's impact on model's responses (Elith et al., 2008; Assis et al., 2022). Ensemble modelling was performed, with the development of distribution maps per species, by averaging the algorithm's predictive responses and the multiple cross-validation rounds (Araújo & New, 2007). The maps were then reclassified into binary maps representing presence and absence data by applying a minimum training threshold. This threshold was determined based on setting a minimum predicted area, while maintaining a sensitivity of 0.95 or higher (Vignali et al., 2020). To minimize overprediction, the maps were then clipped to include areas potentially reachable by dispersal, a crucial step in SDM for low dispersal species such as sponges (Ekins et al., 2016). Therefore, while dispersing, sexual or asexual propagules cannot traverse regions of unsuitable habitats unless supported by documented occurrence records. With this assumption, we set a range for up to 200 km.

Species richness patterns were estimated by summing the individual species' binary maps while considering dispersal constraints (Fragkopoulou et al., 2022). Estimate uncertainty was determined by mapping the standard deviation of the predictive responses the three algorithms and cross-validation rounds. For visualization purpose, the maps were aggregated into 60 km resolution hexagons, and summary tables were generated to reveal habitat coverage and diversity per marine ecoregion (Spalding et al., 2007).

Species presence was denoted in a binary format, with 1 indicating presence and 0 indicating absence, within the Marine Ecoregions of the World's biogeographic classification of global coasts and shelves. This global marine classification system comprises three levels: Marine Realms, Ecoregions and Provinces (Spalding et al., 2007). Hierarchical clustering analysis was performed for the existing 12 marine realms, based on the predicted presence/absence of each species, employing the group-average linking of Bray-Curtis similarities as followed by van Soest et al., 2012. To determine the optimal number of clusters, the silhouette method was used (Rousseeuw, 1987).

All analyses were performed in R (R Core Team, 2022).

3.5 Results

Species distribution models were produced for 649 sponges of class Demospongiae. These species belong to 201 genera, 77 families and 20 orders. In general, the ensemble of the three algorithms (BRT, ADABoost, XGBoost) demonstrated excellent performance when predicting the occurrence of the demosponge species (AUC: 0.99 ± 0.03; TSS: 0.98 ± 0.06; Boyce: 0.90 ± 0.08) (Figure 3.5.1). The application of dispersal constraints marginally improved the metrics' performance (ΔAUC: 0 ± 0.01; ΔTSS: 0± 0.01) or not at all (i.e, ΔBoyce: 0 ± 0) (Figure 3.5.1).

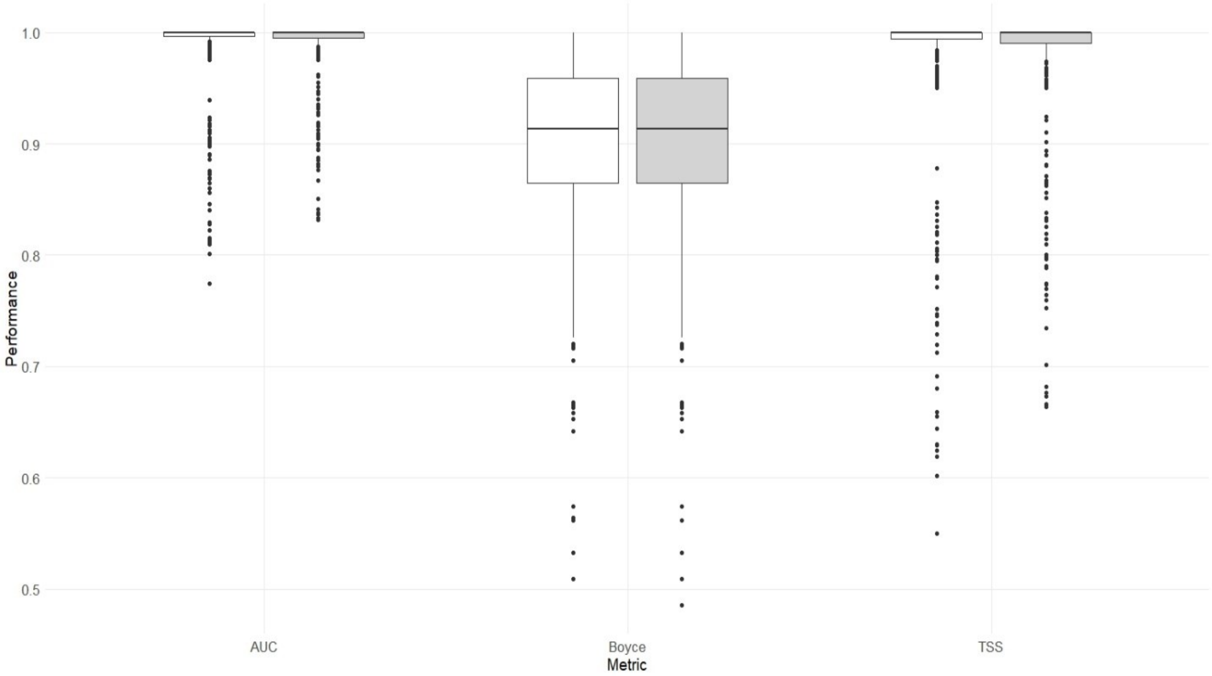


Figure 3.5.1. Performance metrics for the model ensemble of the three algorithms (BRT, ADABoost, XGBoost) estimated for the 649 sponge species: Boyce index, area under the curve (AUC) and true skill statistic (TSS). White boxplots: Before dispersal constraints (left). Gray boxplots: After dispersal constraints (right).

Minimum and maximum temperature, pH and phytoplankton productivity contributed the most to expain the distribution models of demosponges (average relative contribution of 26.98 %, 10.64%, 14.34% and 12.34% respectively). Minimum salinity and mean terrain ruggedness had the lowest contribution to the models, with each accounting for less than 5% (Figure 3.5.2). The models predicted thermal thresholds defining the distribution of demosponges to range between 5.56 ± 9.14 ºC and 27.14 ± 5.93 ºC. Furthermore, overall distributions were predicted in habitats where pH greater than 7.71 ± 0.23 and phytoplankton productivity above 0.07 ± 0.19 µmol/ m^3 (Figure 3.5.3).

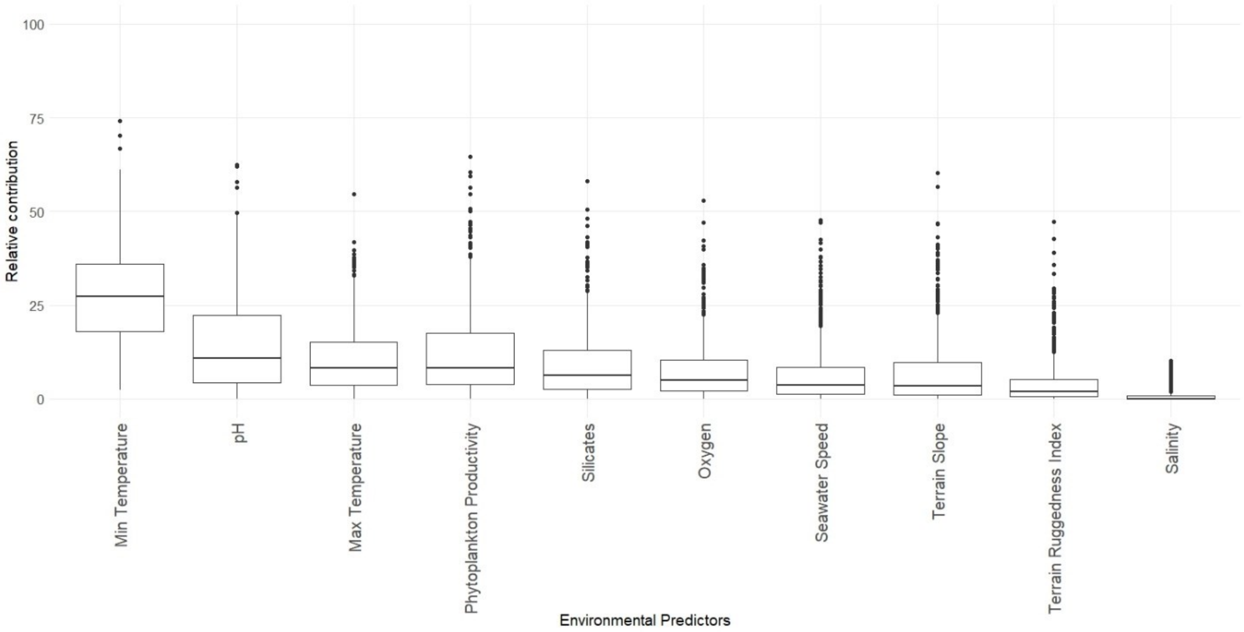


Figure 3.5.2. Relative contribution of the environmental predictors estimated using the ensemble of the models (BRT, ADABoost, XGBoost) for the 649 sponge species.

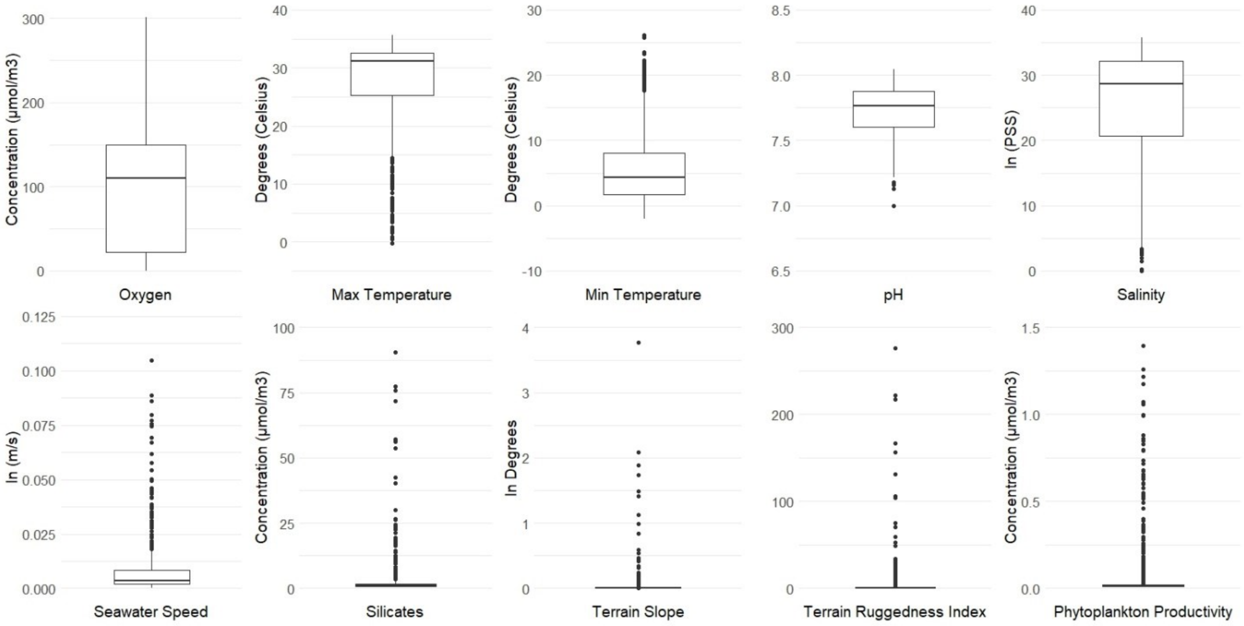


Figure 3.5.3. Tipping points of each environmental predictor as inferred from the ensemble of the models (BRT, ADABoost, XGBoost) for the 649 sponge species.

Stacking individual SDMs enabled the identification of biogeographic patterns of demosponges at a global level (Figure 3.5.4) with low uncertainty (Figure 3.5.5). Overall, suitable habitats for demosponges cover an area of 47,269,253 km^2, distributed at an average depth of 103.83 ± 239.24 meters, with a maximum depth at 8,758 meters for the species *Chondrocladia (Chondrocladia) concrescens* (Supplementary material S1).

In total, model uncertainty was low (maximum: 0.07), detected mostly in regions along the Western Central Atlantic, and the Northeast Pacific (Figure 3.5.5).

One major diversity hotspot was found in the Tropical Atlantic and Temperate Northern Atlantic marine realms (Figures 3.5.4), comprising maximum species richness levels of 245 and 286 species, and an endemicity of 30 and 39 species, respectively (Figure 3.5.7). Specifically, maximum richness was predicted in the Floridian Ecoregion (171 species; Supplementary material S2) followed by the Ecoregion of Northern Gulf of Mexico (66 species; Supplementary material S2). A second hotspot was estimated in the Central Indo-Pacific realm, with maximum richness of 290 species and endemicity of 11 (Figure 3.5.7). Maximum richness was predicted in the Central and Southern Great Barrier Reef (137 species; Supplementary material S2) and Fiji Islands (137 species; Supplementary material S2). Areas with high biodiversity were predicted also in the Gulf of Papua (maximum: 109 species; Supplementary material S2) and Solomon Sea (maximum: 104 species; Supplementary material S2). Additionally, the Eastern Indo-Pacific realm was estimated to have a total maximum richness of 282 species, but no endemics. More precisely, maximum richness was predicted in Gilbert and Ellis Islands (130 species; Supplementary material S2).

Lowest species richness was predicted in Tropical Eastern Pacific and Temperate Southern Africa, comprising 16 and 11 species respectively with no endemics present (3.5.7).

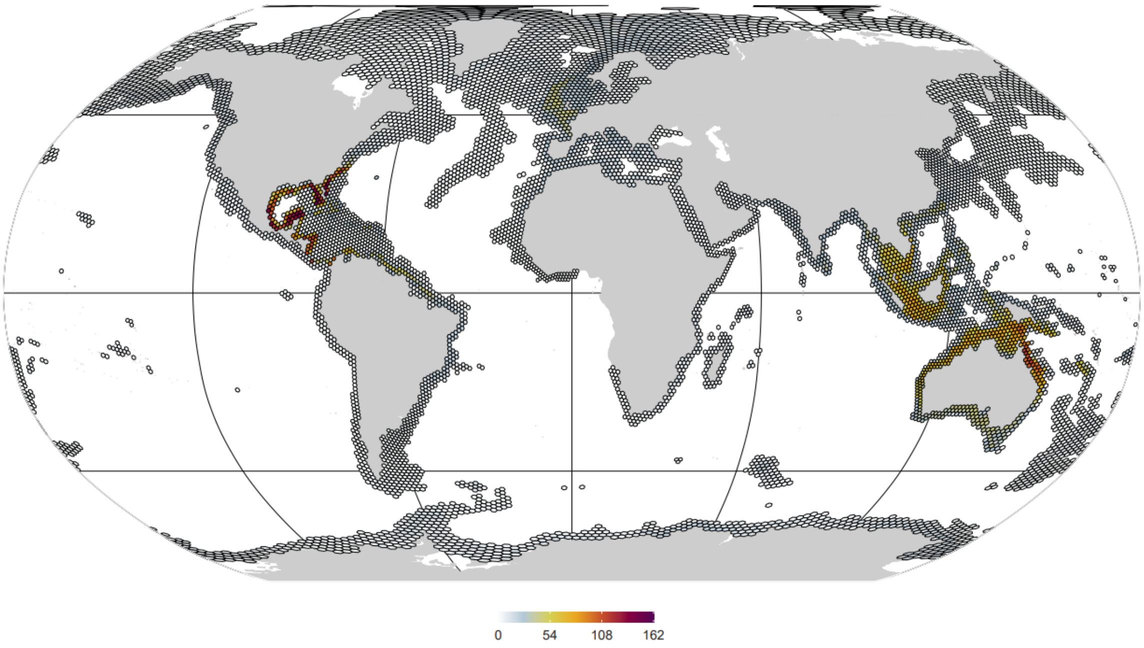


Figure 3.5.4. Global estimates of demosponge species richness predicted with Species Distribution Modelling. The ensemble of the algorithms (BRT, ADABoost, XGBoost) for the individual models of the 649 demosponges was stacked to unravel their patterns on a global scale. For visualization purposes, an optimal resolution of the global hexagon grid system (60 km) was used. The spectrum (0 – 162) indicates the species number estimated within each hexagon.

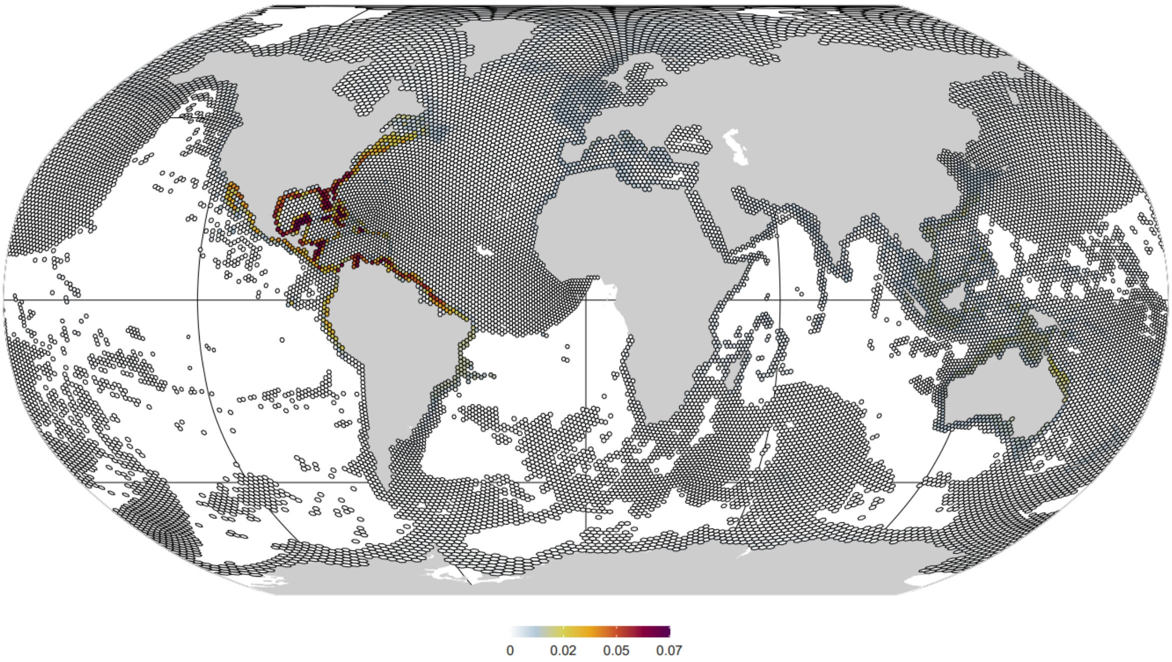


Figure 3.5.5. Uncertainty of the global estimates of demosponge species richness predicted with Species Distribution Modelling. The uncertainty is calculated as the standard deviation of the predictive responses of the three algorithms (BRT, ADABoost, XGBoost) and cross-validation rounds. For visualization purposes, an optimal resolution of the global hexagon grid system (60 km) was used. The spectrum (0 – 0.07) indicates the uncertainty estimated within each hexagon.

The hierarchical clustering of the modelled distribution of species into the twelve marine realms revealed three broad assemblages, supported by the silhouette method (Figure 3.5.6, 3.5.7). The Southern Ocean stood out as a single cluster, having the highest number of endemic species (49). The second cluster consisted of Temperate South America and Tropical Eastern Pacific. Finally, the biggest third cluster comprised geographically neighbouring realms; the Arctic with the Temperate and Tropical Atlantic and the remaining realms (Figure 3.5.7).

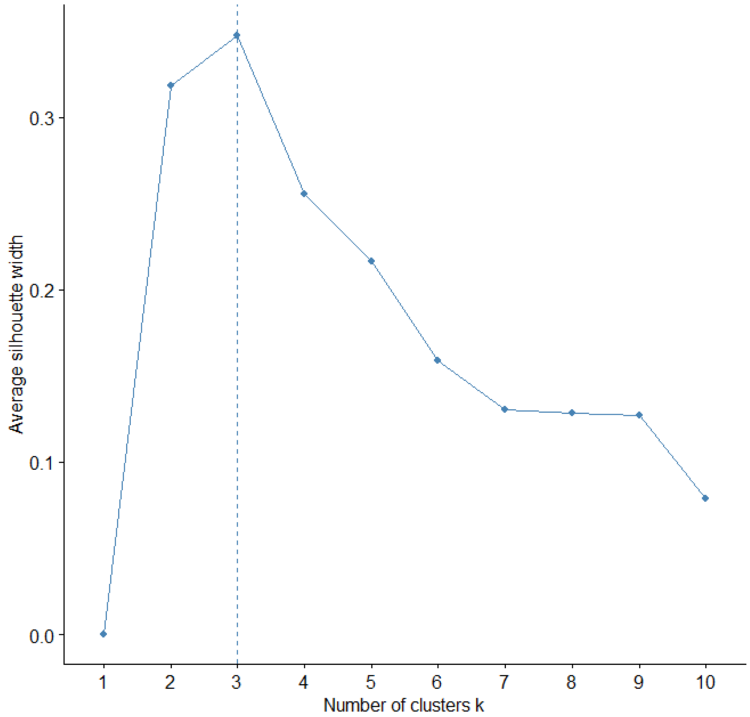
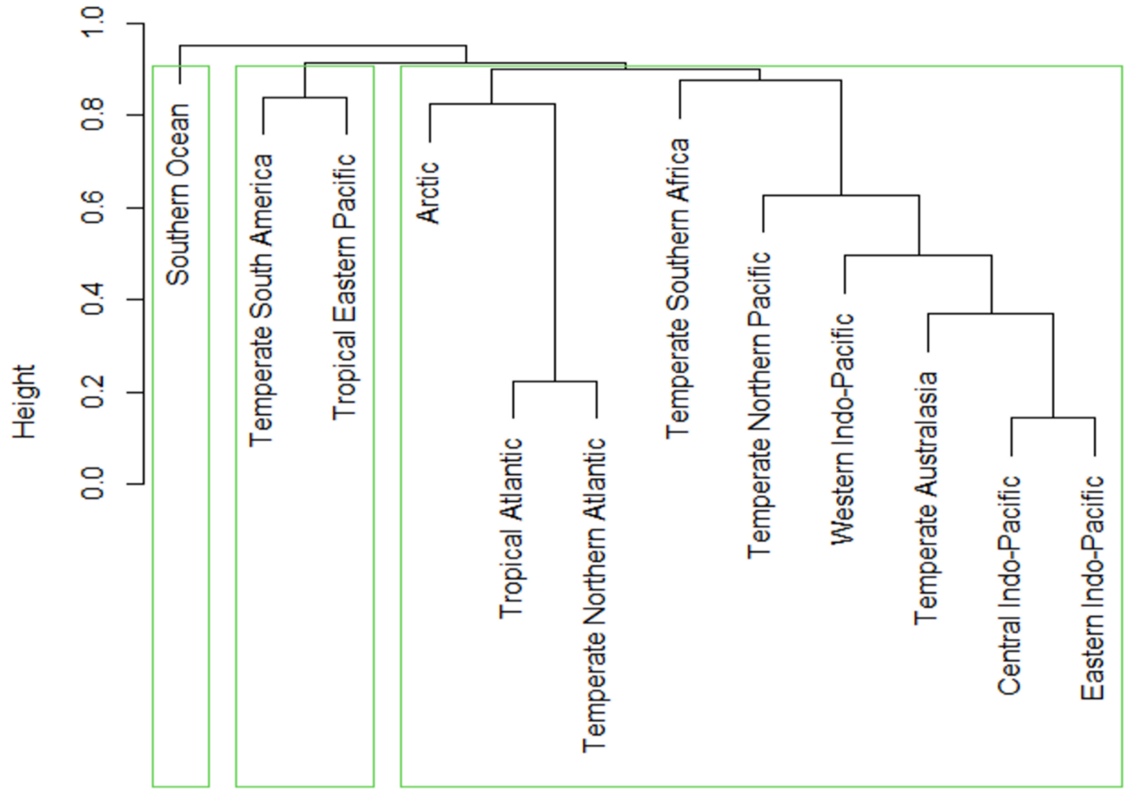


Figure 3.5.6. Silhouette method employed to determine the optimal number of spatial clusters of demosponge species speard across the 12 marine realms, based on presence/absence data of the 649 modelled species.



Richness	72	34	16	60	245	286	11	96	102	159	290	282
Endemicity	49	0	0	0	30	39	5	10	0	4	11	0

Figure 3.5.7. Bray-Curtis dissimilarity dendrogram with average-based linkage method based on predicted presence/absence data of demosponges across the 12 marine realms. Green boxes indicate the spatial clusters defined after the silhouette method. Below bars show species richness and endemicity levels for each realm.

3.6 Discussion

Using stacked species distribution models, we provide the first overall view of the global biodiversity patterns of demosponge species, marking an advancement in our understanding of this important taxonomic group. According to our estimates, the demosponge biome is mostly defined by thermal conditions, while primary productivity and pH are also major distribution drivers. Areas with the highest diversity were predicted in the tropic and subtropic zones. Specifically, key areas of high diversity are the Caribbean Sea and the Gulf of Mexico in Atlantic Ocean and the Coral Triangle and Great Barrier Reef regions in the Indo-Pacific Oceans. Our cluster analysis based on marine realms revealed the existence of three differentiated groups. The Southern Ocean stood alone as a single cluster, showing the highest levels of endemicity, while the Eastern Tropic Pacific assembled with Temperate South America. The remaining marine realms coalesce into a third cluster, implying that the Atlantic together with Indo-Pacific and Australasia host a substantial number of shared species with widespread distributions. Overall, these estimates serve as baselines that can now inform conservation and management strategies, including actions aimed at safeguarding the numerous ecological contributions.

We utilized an ensemble of three machine learning algorithms, an approach used for enhancing the robustness of models (Araujo & New, 2007). These models, trained on expert-validated data, along with a collection of biologically important environmental drivers, exhibited both high performance and low uncertainty. The restricted SDMs provide a more precise representation of the geographical and environmental niche a taxon is exposed to, without needing previous knowledge about species dispersal ecology (Cooper & Soberon, 2018). However, the incorporation of dispersal limitations, a valuable yet underestimated strategy, that generally enhances model performance, did not significantly affected the performance of our models, but allowed controlling overprediction (Mendes et al., 2020).

Our findings support that temperature, pH and primary productivity are the most significant environmental drivers that shape the distribution of the demosponge biome on a global level. Temperature conditions, in particular, were the main driver of distribution patterns, which is consistent with previous studies focused on sponges (Huang et al., 2011; Howel et al., 2016) or in marine biodiversity studies at macroecological scales (Tittensor et al., 2010; Gagné et al., 2020). In addition, pH was inferred to drive demosponge distributions. Sponges, as lower invertebrates, exhibit limited capacities to regulate their acid-base balance due to the absence of protective cellular structures found in other animals, making them potentially vulnerable to pH fluctuations (Pörtner, 2008; Goodwin et al., 2014). Phytoplankton productivity was further estimated to be a major contributor in sponge distribution. Indeed, phytoplankton productivity fuels energy flux into marine ecosystems, and sponges being prominent filter-feeders process large quantities of dissolved and particulate organic matter and are primarily mixotrophic or heterotrophic (Valentine & Jablonski, 2019; Lesser & Slattery, 2020; de Goeij et al., 2017). It is worth mentioning that, in addition to their filter-feeding strategy, sponges can also rely on phototrophic symbionts to meet their metabolic and energetic needs or, on rare occasions, exhibit carnivory (Wilkinson & Cheshire, 1990; Vacelet & Boury-Esnault, 1995). Other physical (salinity, silicates, oxygen), disturbance (currents) and topographic predictors (slope, terrain ruggedness) were not as important for all species, yet they were found key for specific species. For instance, for *Isodictya erinacea,* the relative contribution of silicates was 58.10%, for *Hymeniacidon heliophile* salinity's relative contribution was 10.14%, and for *Spirastrella hartmani,* terrain ruggedness dominated the relative contribution of predictors, on a 69.52%. Altogether, our study has indicated the key environmental predictors shaping the distribution of demosponges along with their overall critical thresholds for the first time. Given the fact that there is a knowledge gap on how most sponges respond to environmental pressures (Bell et al., 2015), these findings can serve as a foundation for future experimental studies. It is necessary to consider that responses to environmental perturbations might exhibit species-specific characteristics and therefore influence variations in resilience beyond these thresholds (Guzman & Conaco, 2016).

Sponges, being major components of benthic communities from shallow to abyssal depths, were predicted to have a vast extent of suitable habitats that reaches c. 47 million km^2 on a global scale. In comparison, other ecosystem structuring species, such as kelps, cover significantly less area, c. 1.71 million km^2 (Fragkopoulou et al., 2022), which is roughly 27 times smaller than our predicted estimate for sponges. Others, such as mangroves or zooxanthellate corals, exhibit coverage areas more than 300 times smaller (Jayathilake & Costello, 2020). These are coastal, light-dependent taxa, restricted to a narrow vertical distribution, while sponges have the advantage of thriving in a variety of environments not dependent on light availability (Amina & Musayeib, 2018). While our finding constitutes a first worldwide estimate, future research should use higher spatial resolution to identify patchy distributions of demosponges at the regional scale.

Our stacked species distribution models allowed us to discern the distribution patterns of the demosponges on a global level. A marine latitudinal diversity gradient was revealed, with high species richness existing in low latitudes (Supplementary material S3). This observation aligns with previous macroecological studies made across taxa that showed a similar trend of diversity reaching its peak in low latitudes and declining poleward (Valentine & Jablonski, 2015; Gagné et al., 2017). Likewise, two major hotspots in these low latitudes were predicted; one located in the Western Atlantic, covering the Caribbean Sea and the Gulf of Mexico, and the second in the Indo-Pacific, encompassing the Coral Triangle area and Great Barrier Reef in the North-East Australia. These results coincide with the two primary coral reef areas that occur in the global ocean, the "Greater Caribbean" which extends from Bermuda down to Venezuela and the second one in the Indo-Pacific (Maldonado et al., 2017). Especially, the Coral Triangle is traditionally considered to host the greatest species diversity in the world (Brigss & Bowen, 2013), which our study and additional ones show that it is also implied for sponges (Maldonado et al., 2017), while the Caribbean hosting the greatest concentration of marine species in the Atlantic Ocean (Roberts et al., 2002). A few evolutionary theories may explain the observed high diversity in these two regions (Bowen et al., 2013). One is the center of speciation or origin, where these epicenters act as species exporters potentially following the population fragmentation caused by complex long-term geological features and habitat diversity, coupled with competition. Additionally, the center of accumulation theory suggests that speciation occurs in peripheral habitats which then disperse towards the center, favored by oceanographic currents. The center of overlap is another theory, developed only for the Indo-Pacific region, which supports that diversity stems from distinct Indian and Pacific faunas developed during periods of isolation during low sea levels. A more recent theory, biodiversity feedback, includes elements of all previous models, suggesting allopatric and sympatric speciation concurs in the marine environment in both directions enhancing biodiversity. (Bowen et al., 2013)

Within the tropics, differences in diversity have also been inferred regarding our results between the western and eastern sides of the Atlantic and the Pacific Ocean, with the eastern parts appearing less diverse. This disparity may be attributed to equatorial ocean currents which are directed towards higher latitudes in the western sides of the two oceans, carrying warm water and tropical species with them, while in the eastern sides bring cool waters towards the tropics, resulting in lower diversity in the latter areas (Briggs, 2003). The Eastern Tropical Pacific (ETP), being a small marine realm, is considered an isolated oceanic region exposed to extreme oceanographic conditions whose coral reefs possess a less extensive suite of species in comparison with other coral reef areas (Cortés et al., 2017). This results in its sponge fauna being less diverse than those of Caribbean and Western Pacific reefs. In ETP, sponges are mainly cryptic encrusting and very small in size (Carballo et al., 2019). In the case of the respective eastern side of Tropical Atlantic, which is considered a subset of the Tropical Atlantic marine realm, studies regarding sponge diversity are very limited (Bell et al., 2020).

Our cluster analysis based on species presence and absence data revealed three main groups within twelve marine realms. The widest cluster comprises all major marine realms, among which we detected the biodiversity hotspot areas. This suggests that there are shared species among the Indo-Pacific Ocean and the Atlantic, which might have exploited dispersal passages and opportunities across geological time scales. For instance, the Cape of Good Hope in South Africa intermittently allowed westward tropical species colonization from the Western Indian Ocean to the Atlantic (Briggs & Bowen, 2013). Another association resides in the north, where the presence of modern sponge species pairs/assemblages in the Arctic and the adjacent regions indicate faunal connections, that have might been evolved from the Pacific (Morozov et al., 2021). Furthermore, the Suez Canal's 1869 construction facilitated the entry of Indo-Pacific biota into the Mediterranean, with reports of sponges originating from the Red Sea dated as early as 1936 (Galil & Zenetos, 2002; Tsurnamal, 1969). The clustering results further put the Southern Ocean in a single cluster and with notable endemism. This observation aligns with prior studies that consider the Antarctic demosponge biome a single distinct fauna (Downey et al., 2012). The Polar front, linked with the Antarctic Circumpolar Current, serves as a dispersal barrier both to and from other regions of lower Southern latitudes (Griffiths et al., 2009; Downey et al., 2012). However, despite this barrier, it is worth noting that faunal associations exist between Antarctic and South America both in different taxa, such as bivalves (Griffiths et al., 2009) and also sponges (Downey et al., 2012). The third cluster showed stronger biogeographic affinities of Temperate South America with Eastern Tropic Pacific. In a broader context of the ETP fauna, it has been observed that affinities vary among taxa, with some out of the well-studied, like fishes, showing strong connections to the Indo-Pacific, but this can be taxa-dependent (Cortés et al., 2017). Nevertheless, research on ETP sponge distribution is scarce, preventing from revealing any biogeographic affinities of the resident sponges (Vega et al., 2012).

While our models exhibit high performance, it is essential to address some inherent limitations of our methods. Generally, there is likely an uneven research effort, considering the research hotspot of the North Atlantic Ocean (Melo-Merino et al., 2020). In particular for the phylum Porifera, the Temperate Northern Atlantic dominates literature (40% of collected studies; Bell et al., 2015). Such an imbalance could potentially affect global biodiversity estimates. Furthermore, errors in taxonomic classification of sponges can also impair results. Sponge diversity has been greatly underestimated (Knowlton, 2000; Wörheide et al., 2005), since previously misidentified cryptic species are now being discovered using genetic markers (Uriz & Turon, 2012). Such a case is *Cliona celata*, which is considered a cosmopolitan species, but a study showed that it constitutes a species complex comprised of at least four morphologically indistinct species with more restricted distributions (Xavier et al., 2010). Additionally, while abiotic factors play a crucial role in determining sponge presence, ecological interactions can further enhance or restrict the impact of these factors (Wulff, 2012). Gaps in this type of information together with factors such as substrate habitat on a global level, which has been found to be of major importance (Barnes & Bell, 2002) could help us improve our models.

To our knowledge, this research provides the first estimate of demosponge marine biodiversity on a global scale, on the basis of Species Distribution Modelling. After negotiating and developing the post-2020 global biodiversity framework, parties to the Convention on Biological Diversity have now set up goals and targets for positive biodiversity change up to 2030. Given the diverse ecological roles of sponges in marine ecosystems, our estimates serve now as a baseline to meet such conservation and management goals (Sofaer et al., 2019). Our results can guide prospective sampling surveys in a cost-effective way to avoid mismatch between biodiversity and conservation spending (Halpern et al., 2006). Areas of high diversity should take precedence. Moreover, regions with estimated thriving populations but limited research, such as the South Gulf of Mexico require special emphasis (Ugalde et al., 2015). Additionally, areas with insufficient data like the Eastern Tropical Atlantic or zones which exhibit high endemicity levels such as the Southern Ocean, also demand focused attention (Bell et al., 2020). Demosponges are essential benthic components of marine environment contributing significantly to the health, function and services of marine ecosystems, therefore their understanding and protection will benefit ecosystem health and human well-being.

Ethics statements

The present work complies with ethical requirements and does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT author statement

Conceptualization: VA, EF, and JA. Data curation: VA and JA. Funding acquisition: JA and ES. VA led the writing with support from JA and EF. All authors contributed to editing the article and approved the submitted version.

Acknowledgments

This study received Portuguese national funds from the Horizon Europe Framework Programme through project MPAEurope (HORIZON-CL6-2021-BIODIV-01-12) and the Foundation for Science and Technology (FCT) through programs UIDB/04326/2020, UIDP/04326/2020, LA/P/0101/2020 and PTDC/BIA-CBI/6515/2020, the Individual Call to Scientific Employment Stimulus 2022.00861 to J.A. and the fellowship SFRH/BD/144878/2019 to E.F

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

3.7 References

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of applied ecology, 43(6), 1223-1232.

Amina, M., & Al Musayeib, N. M. (2018). Biological and medicinal importance of sponge. Biological Resources of Water.

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in ecology & evolution, 22(1), 42-47.

Assis, J., Berecibar, E., Claro, B., Alberto, F., Reed, D., Raimondi, P., & Serrão, E. A. (2017a). Major shifts at the range edge of marine forests: the combined effects of climate changes and limited dispersal. Scientific Reports, 7(1), 44348.

Assis, J., Castilho Coelho, N., Alberto, F., Valero, M., & Raimondi, P. (2013). High and Distinct Range-Edge Genetic Diversity despite Local Bottlenecks. PLoS ONE, 8(7), e68646.

Assis, J., Serrão, E. A., Duarte, C. M., Fragkopoulou, E., & Krause-Jensen, D. (2022). Major expansion of marine forests in a warmer Arctic. Frontiers in Marine Science, 9, 850368.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., De Clerck, O., & Tittensor, D. (2017b). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. Global Ecology and Biogeography, 27(3), 277-284.

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(2), 327-338.

Barnes, D. K. (1999). High diversity of tropical intertidal zone sponges in temperature, salinity and current extremes. African Journal of Ecology, 37(4), 424-434.

Barnes, D. K., & Bell, J. J. (2002). Coastal sponge communities of the West Indian Ocean: taxonomic affinities, richness and diversity. African Journal of Ecology, 40(4), 337-349.

Bell, J. J. (2008). The functional roles of marine sponges. Estuarine, coastal and shelf science, 79(3), 341-353.

Bell, J. J., Bennett, H. M., Rovellini, A., & Webster, N. S. (2018). Sponges to be winners under near-future climate scenarios. Bioscience, 68(12), 955-968.

Bell, J. J., McGrath, E., Biggerstaff, A., Bates, T., Cárdenas, C. A., & Bennett, H. (2015). Global conservation status of sponges. Conservation Biology, 29(1), 42-53.

Bell, J. J., McGrath, E., Kandler, N. M., Marlow, J., Beepat, S. S., Bachtiar, R., ... & Carballo, J. L. (2020). Interocean patterns in shallow water sponge assemblage structure and function. Biological Reviews, 95(6), 1720-1758.

Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M., & Courchamp, F. (2013). Will climate change promote future invasions?. Global change biology, 19(12), 3740-3748.

Bowen, B. W., Rocha, L. A., Toonen, R. J., & Karl, S. A. (2013). The origins of tropical marine biodiversity. Trends in ecology & evolution, 28(6), 359-366.

Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. (2002). Evaluating resource selection functions. Ecological modelling, 157(2-3), 281-300.

Briggs, J. C. (2003). Guest editorial: marine centres of origin as evolutionary engines. Journal of Biogeography, 30(1), 1-18.

Briggs, J. C., & Bowen, B. W. (2013). Marine shelf habitat: biogeography and evolution. Journal of Biogeography, 40(6), 1023-1035.

Carballo, J. L., Bautista-Guerrero, E., & Leyte-Morales, G. E. (2008). Boring sponges and the modeling of coral reefs in the east Pacific Ocean. Marine Ecology Progress Series, 356, 113-122.

Carballo, J. L., Cruz-Barraza, J. A., Vega, C., Nava, H., & Chávez-Fuentes, M. D. C. (2019). Sponge diversity in Eastern Tropical Pacific coral reefs: an interoceanic comparison. Scientific reports, 9(1), 9409.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Citores, L., Ibaibarriaga, L., Lee, D. J., Brewer, M. J., Santos, M., & Chust, G. (2020). Modelling species presence–absence in the ecological niche theory framework using shape-constrained generalized additive models. Ecological Modelling, 418, 108926.

Cooper, J. C., & Soberón, J. (2018). Creating individual accessible area hypotheses improves stacked species distribution model performance. Global Ecology and Biogeography, 27(1), 156-165.

Cortés, J., Enochs, I. C., Sibaja-Cordero, J., Hernández, L., Alvarado, J. J., Breedy, O., ... & Zapata, F. A. (2017). Marine biodiversity of Eastern Tropical Pacific coral reefs. Coral reefs of the Eastern Tropical Pacific: persistence and loss in a dynamic environment, 203-250.

de Goeij, J. M., Lesser, M. P., & Pawlik, J. R. (2017). Nutrient fluxes and ecological functions of coral reef sponges in a changing ocean. Climate change, ocean acidification and sponges: impacts across multiple levels of organization, 373-410.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., d'Amen, M., Randin, C., ... & Guisan, A. (2017). ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography, 40(6), 774-787.

Diaz, M. C., & Rützler, K. (2001). Sponges: an essential component of Caribbean coral reefs. Bulletin of marine science, 69(2), 535-546.

Downey, R. V., Griffiths, H. J., Linse, K., & Janussen, D. (2012). Diversity and distribution patterns in high southern latitude sponges. PLoS One, 7(7), e41672.

Ekins, M., Erpenbeck, D., Wörheide, G., & Hooper, J. N. (2016). Staying well connected–Lithistid sponges on seamounts. Journal of the Marine Biological Association of the United Kingdom, 96(2), 437-451.

Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., ... & E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29(2), 129-151.

Elith, J., & Graham, C. H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography, 32(1), 66-77.

Elith, J., & Leathwick, J. (2017). Boosted Regression Trees for ecological modeling. R Documentation. Available online: https://cran. r-project. org/web/packages/dismo/vignettes/brt. pdf (accessed on 12 June 2011).

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. Journal of animal ecology, 77(4), 802-813.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30(5), 609-628.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation, 24(1), 38-49.

Fragkopoulou, E., Serrão, E. A., De Clerck, O., Costello, M. J., Araújo, M. B., Duarte, C. M., ... & Assis, J. (2022). Global biodiversity patterns of marine forests of brown macroalgae. Global Ecology and Biogeography, 31(4), 636-648.

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. Diversity and distributions, 19(10), 1217-1223.

Gagné, T. O., Reygondeau, G., Jenkins, C. N., Sexton, J. O., Bograd, S. J., Hazen, E. L., & Van Houtan, K. S. (2020). Towards a global understanding of the drivers of marine and terrestrial biodiversity. PLoS One, 15(2), e0228065.

Galil, B. S., & Zenetos, A. (2002). A sea change—exotics in the Eastern Mediterranean Sea. Invasive aquatic species of Europe. Distribution, impacts and management, 325-336.

Goodwin, C., Rodolfo-Metalpa, R., Picton, B., & Hall-Spencer, J. M. (2014). Effects of ocean acidification on sponge communities. Marine Ecology, 35, 41-49.

Griffiths, H. J., Barnes, D. K., & Linse, K. (2009). Towards a generalized biogeography of the Southern Ocean benthos. Journal of Biogeography, 36(1), 162-177.

Guzman, C., & Conaco, C. (2016). Gene expression dynamics accompanying the sponge thermal stress response. PloS one, 11(10), e0165368.

Halpern, B. S., Pyke, C. R., Fox, H. E., Chris Haney, J., Schlaepfer, M. A., & Zaradic, P. (2006). Gaps and mismatches between global conservation priorities and spending. Conservation Biology, 20(1), 56-64.

Heikkinen, R. K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability?. Ecography, 35(3), 276-288.

Hofner, B., Müller, J., & Hothorn, T. (2011). Monotonicity-constrained species distribution models. Ecology, 92(10), 1895-1901.

Howell, K. L., Piechaud, N., Downie, A. L., & Kenny, A. (2016). The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management. Deep Sea Research Part I: Oceanographic Research Papers, 115, 309-320.

Huang, Z., Brooke, B., & Li, J. (2011). Performance of predictive models in marine benthic environments based on predictions of sponge distribution on the Australian continental shelf. Ecological Informatics, 6(3-4), 205-216.

Idan, T., Shefer, S., Feldstein, T., Yahel, R., Huchon, D., & Ilan, M. (2018). Shedding light on an East-Mediterranean mesophotic sponge ground community and the regional sponge fauna. Mediterranean Marine Science, 19(1), 84-106.

IUCN. (2016). Increasing marine protected area coverage for effective marine biodiversity conservation. In WCC-2016-Res-050-EN, World Conservation Congress, Hawai ‘i, United States of America, 1–10 September 2016.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent data analysis, 6(5), 429-449.

Jayathilake, D. R., & Costello, M. J. (2020). A modelled global distribution of the kelp biome. Biological Conservation, 252, 108815.

Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and distributions, 14(6), 885-890.

Knowlton, N. (2000). Molecular genetic analyses of species boundaries in the sea. Marine genetics, 73-90.

Krause-Jensen, D., Archambault, P., Assis, J., Bartsch, I., Bischof, K., Filbee-Dexter, K., ... & Duarte, C. M. (2020). Imprint of climate change on pan-Arctic marine vegetation. Frontiers in Marine Science, 7, 617324.

Lesser, M. P., & Slattery, M. (2020). Will coral reef sponges be winners in the Anthropocene?. Global Change Biology, 26(6), 3202-3211.

Maldonado, M., Aguilar, R., Bannister, R., Bell, J., Conway, J., Dayton, P., ... & Young, C. (2017). Sponge grounds as key marine habitats: a synthetic review of types, structure, functional roles, and conservation concerns. Marine animal forests: The ecology of benthic biodiversity hotspots.

Mehbub, M. F., Lei, J., Franco, C., & Zhang, W. (2014). Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. Marine drugs, 12(8), 4539-4577.

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. Ecological Modelling, 415, 108837.

Mendes, P., Velazco, S. J. E., de Andrade, A. F. A., & Júnior, P. D. M. (2020). Dealing with overprediction in species distribution models: How adding distance constraints can improve model accuracy. Ecological Modelling, 431, 109180.

Morozov, G., Sabirov, R., & Anisimova, N. (2021). The hidden diversity of the endemic Arctic sponges (Porifera). Journal of Natural History, 55(9-10), 571-596.

Morrow, C., & Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). Frontiers in Zoology, 12, 1-27.

Muir, L. A., Botting, J. P., & Beresi, M. S. (2017). Lessons from the past: Sponges and the geological record. Climate Change, Ocean Acidification and Sponges: Impacts Across Multiple Levels of Organization, 13-47.

Pomponi, S. A., Diaz, M. C., Van Soest, R. W., Bell, L. J., Busutil, L., Gochfeld, D. J., ... & Slattery, M. (2019). Sponges. Mesophotic coral ecosystems, 563-588.

Pörtner, H. O. (2008). Ecosystem effects of ocean acidification in times of ocean warming: a physiologist's view. Marine Ecology Progress Series, 373, 203-217.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roberts, C. M., McClean, C. J., Veron, J. E., Hawkins, J. P., Allen, G. R., McAllister, D. E., ... & Werner, T. B. (2002). Marine biodiversity hotspots and conservation priorities for tropical reefs. Science, 295(5558), 1280-1284.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.

Segurado, P. A. G. E., Araujo, M. B., & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology, 43(3), 433-444.

Senay, S. D., Worner, S. P., & Ikeda, T. (2013). Novel three-step pseudo-absence selection technique for improved species distribution modelling. PloS one, 8(8), e71218.

Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., ... & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. BioScience, 69(7), 544-557.

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M. A. X., ... & Robertson, J. (2007). Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. BioScience, 57(7), 573-583.

Tiller, R., & Mendenhall, E. (2023). And so it begins–The adoption of the 'Biodiversity Beyond National Jurisdiction'treaty. Marine Policy, 157, 105836.

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. Nature, 466(7310), 1098-1101.

Tsurnamal, M. (1969). Sponges of Red Sea origin on the Mediterranean coast of Israel. Israel Journal of Zoology, 18(2-3), 149-155.

Ugalde, D., Gómez, P., & Simões, N. (2015). Marine sponges (Porifera: Demospongiae) from the Gulf of México, new records and redescription of Erylus trisphaerus (de Laubenfels, 1953). Zootaxa, 3911(2), 151-183.

Uriz, M. J., & Turon, X. (2012). Sponge ecology in the molecular era. Advances in marine biology, 61, 345-410.

Vacelet, J., & Boury-Esnault, N. (1995). Carnivorous sponges. Nature, 373(6512), 333-335.

Vad, J., Barnhill, K. A., Kazanidis, G., & Roberts, J. M. (2021). Human impacts on deep-sea sponge grounds: Applying environmental omics to monitoring. In Advances in Marine Biology (Vol. 89, pp. 53-78). Academic Press.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs, 92(1), e01486.

Valavi, R., Shafizadeh-Moghadam, H., Matkan, A., Shakiba, A., Mirbagheri, B., & Kia, S. H. (2019). Modelling climate change effects on Zagros forests in Iran using individual and ensemble forecasting approaches. Theoretical and Applied Climatology, 137, 1015-1025.

Valentine, J. W., & Jablonski, D. (2015). A twofold role for global energy gradients in marine biodiversity trends. Journal of Biogeography, 42(6), 997-1005.

van Proosdij, A. S., Sosef, M. S., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. Ecography, 39(6), 542-552.

Van Soest, R. W., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., De Voogd, N. J., ... & Hooper, J. N. (2012). Global diversity of sponges (Porifera). PLoS one, 7(4), e35105.

Vega, C., Hernández Guerrero, C. J., & Cruz Barraza, J. A. (2012). Biogeografía de esponjas marinas (Phylum Porifera): estudios en el Pacífico Oriental.

Vignali, S., Barras, A. G., Arlettaz, R., & Braunisch, V. (2020). SDMtune: An R package to tune and evaluate species distribution models. Ecology and Evolution, 10(20), 11488-11506.

Wilkinson, C. R., & Cheshire, A. C. (1990). Comparisons of sponge populations across the Barrier Reefs of Australia and Belize: evidence for higher productivity in the Caribbean. Marine ecology progress series. Oldendorf, 67(3), 285-294.

Wörheide, G., Solé-Cava, A. M., & Hooper, J. N. (2005). Biodiversity, molecular ecology and phylogeography of marine sponges: patterns, implications and outlooks. Integrative and comparative biology, 45(2), 377-385.

WoRMS Editorial Board. (2023). World Register of Marine Species. VLIZ. https://www.marinespecies.org. doi:10.14284/17

Wulff, J. (2012). Ecological interactions and the distribution, abundance, and diversity of sponges. Advances in Marine Biology, 61, 273-344.

Xavier, J. R., Rachello-Dolmen, P. G., Parra-Velandia, F., Schönberg, C. H. L., Breeuwer, J. A. J., & Van Soest, R. W. M. (2010). Molecular evidence of cryptic speciation in the "cosmopolitan" excavating sponge Cliona celata (Porifera, Clionaidae). Molecular phylogenetics and evolution, 56(1), 13-20.

Supplementary Material

S1 Estimated minimum, mean and maximum depth per species (provided in a separate excel file; Supplementary material, sheet S1)

S2 Estimated species richness for Marine Realms, Ecoregions and Provinces (provided in a separate excel file; Supplementary material, sheet S2)

S3 Estimated average species number of demosponges in regard to latitude

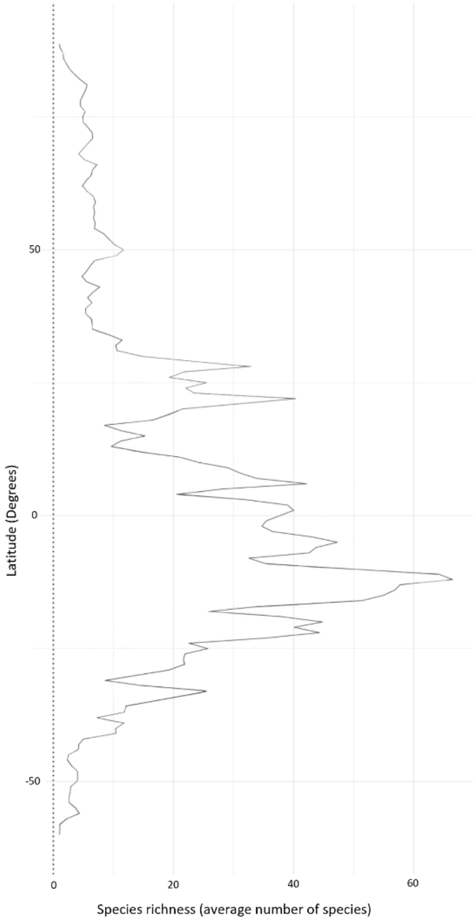


Figure S3.1. Estimated average species number of sponges in regard to latitude.