

# Ecological Informatics

## Spatial and temporal representation of marine fish occurrences available online

--Manuscript Draft--

<b>Manuscript Number:</b>	ECOINF-D-23-00167
<b>Article Type:</b>	Research Paper
<b>Keywords:</b>	Species Richness, Spatial and Temporal Representativeness, Marine Fish, Ecological Information Biases, Ecoinformatics
<b>Corresponding Author:</b>	Horacio Samaniego, PhD Austral University of Chile - Campus Isla Teja CHILE
<b>First Author:</b>	Vanessa Pizarro
<b>Order of Authors:</b>	Vanessa Pizarro Andrea Castillo Andrea Piñones, PhD Horacio Samaniego, PhD
<b>Abstract:</b>	<p>Despite the 243,000 species of marine species described by 2022, our knowledge about the biodiversity in the oceans is still incomplete. This may have dreadful and detrimental effect for the conservation of marine ecosystems under the current anthropization of our biota and the fast pacing climate and global change scenario. However, a large number of online repositories cataloging, storing and distributing biodiversity information hosting taxonomic information and species occurrence data have emerged recently. FishBase, the Global Biodiversity Information Facility (GBIF) and the Ocean Biodiversity Information System (OBIS) are part of these publicly available repositories representing a variety of sources that have exploded in number. However, despite the incredible accumulation of biodiversity records, not all the information is really useful, nor does it represent any new knowledge regarding global species richness patterns.</p> <p>In this study we evaluated the spatial and temporal representativeness of the records of marine fishes (order Actinopterygii) available in the GBIF and OBIS global repositories. We provide a methodological framework based on a set of non-parametric estimators to calculate species richness from incidence data, using hexagonal grids as sampling units overlapped on the marine bioregions worldwide. Using standard ecological and spatial analysis tools, we identify regions that are adequately represented in terms of available records and therefore have more reliable data, as well as regions with few records that do not represent current species richness. We overlap these results with the location of marine protected areas and fishing exploitation zones to understand the anthropogenic effect on marine ichthyofauna. We additionally evaluate hypotheses regarding the taxonomic, geographic, and temporal distribution of information biases to deepen our current understanding of public records of species occurrences worldwide.</p> <p>Considering that more than 40 years of information was analyzed, the results showed that on a global scale, the primary data on marine fish available on GBIF and OBIS platforms are still far from being representative and complete. Only 1.14% of the records were useful for our analyses. In addition, we found that the information seems to be biased towards coastal areas, regions close to developed countries and areas where there is a large fishing activity. Finally, the best represented species and families are those with a small body size, which use shallow habitats and have commercial cultural value.</p>
<b>Suggested Reviewers:</b>	<p>Camilo Mora professor, University of Hawai'i System <a href="mailto:moracamilo@hotmail.com">moracamilo@hotmail.com</a></p> <p>Derek P Tittensor Dalhousie University <a href="mailto:derekt@mathstat.dal.ca">derekt@mathstat.dal.ca</a></p>

Derek Corcoran  
Aarhus University  
[derek.corcoran@bio.au.dk](mailto:derek.corcoran@bio.au.dk)

Signe Normand  
Aarhus University  
[signe.normand@bio.au.dk](mailto:signe.normand@bio.au.dk)

Duccio Rocchini  
University of Bologna  
[duccio.rocchini@unibo.it](mailto:duccio.rocchini@unibo.it)

Horacio Samaniego  
Laboratorio de Ecoinformática  
Universidad Austral de Chile  
Valdivia, Los Ríos, Chile

Valdivia, February 2, 2023

**Subject:** Manuscript Submission

**Journal:** Ecological Informatics

Dear Editor-in-Chief,

We hereby submit a manuscript entitled: "*Spatial and temporal representation of marine fish occurrences available online*" authored by Vanessa Pizarro, Andrea Castillo, Andrea Piñones and myself.

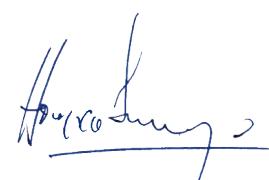
The marine ecosystem is among the least explored of our biota and understanding the spatial distribution of its biodiversity is essential to produce adequate protection planning nowadays. Several ecoinformatics tools have now been developed for such task with sometimes uncertain outcome; vast knowledge is produced at some locations while other are left unsampled. This work tackles precisely on this issue and seeks to provide a description of how marine fish diversity has been collected and stored in publicly available repositories such as OBIS and GBIF. By using a combination of spatial analyses and more classic diversity evaluation, we have identified the current state of the art among fish biodiversity inventories. We identify gaps, associated sample efforts and potential determinants e.g. life history features. Finally, an assessment on the incidence of protection areas and cultural values upon fish species richness is also provided.

We believe that this manuscript is of great interest for your audience as it addresses a unifying theme across biogeography and ecoinformatics by building on prominent contributions in the field, current effort to index our biota, so that we may pinpoint where efforts should be placed for conservation.

We hope that you find the manuscript compelling and would appreciate its prompt evaluation.

Thank you in advance for the consideration,

Sincerely,



A handwritten signature in blue ink, appearing to read "Horacio Samaniego".

horacio@ecoinformatica.cl

# Spatial and temporal representation of marine fish occurrences available online

3      Vanessa Pizarro<sup>a</sup>, Andrea Castillo<sup>a,b</sup>, Andrea Piñones<sup>c,d,e,f</sup>, Horacio Samaniego<sup>a,g,\*</sup>

<sup>a</sup>*Laboratorio de Ecoinformática, Instituto de Conservación, Biodiversidad y Territorio, Universidad Austral de Chile, Valdivia, Chile*

6      <sup>b</sup>*Programa de Doctorado en Ciencias mención Ecología y Evolución, Escuela de Graduados, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile*

9      <sup>c</sup>*Instituto de Ciencias Marinas y Limnológicas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile*

12     <sup>d</sup>*Centro FONDAP de Investigación en Dinámica de Ecosistemas Marinos de Altas Latitudes (IDEAL), Valdivia, Chile.*

15     <sup>e</sup>*Centro de Investigación Oceanográfica COPAS-COASTAL, Universidad de Concepción, Chile*

18     <sup>f</sup>*Millenium Institute Biodiversity of Antarctic and Subantarctic Ecosystems - BASE, Chile*

21     <sup>g</sup>*Instituto de Sistemas Complejos de Valparaíso, Subida Artillería 470, Valparaíso, Chile*

## 15 Abstract

Despite the 243,000 species of marine species described by 2022, our knowledge about the biodiversity in the oceans is still incomplete. This may have dreadful and detrimental effect for the conservation of marine ecosystems under the current anthropization of our biota and the fast pacing climate and global change scenario.

However, a large number of online repositories cataloging, storing and distributing biodiversity information hosting taxonomic information and species occurrence data have emerged recently. FishBase, the Global Biodiversity Information Facility (GBIF) and the Ocean Biodiversity Information System (OBIS) are part of these publicly available repositories representing a variety of sources that have exploded in number. However, despite the incredible accumulation of biodiversity records, not all the information is really useful, nor does it represent any new knowledge regarding global species richness patterns.

In this study we evaluated the spatial and temporal representativeness of the records of marine fishes (order Actinopterygii) available in the GBIF and OBIS global repositories. We provide a methodological framework based on a set of non-parametric estimators to calculate species richness from incidence data, using hexagonal grids as sampling units overlapped on the marine bioregions worldwide.

Using standard ecological and spatial analysis tools, we identify regions that are adequately represented in terms of available records and therefore have more reliable data, as well as regions with few records that do not represent current species richness. We overlap these results with the location of marine protected areas and fishing exploitation zones to understand the anthropogenic effect on marine ichthyofauna. We additionally evaluate hypotheses regarding the taxonomic, geographic, and temporal distribution of information biases to deepen our current understanding of public records of species occurrences worldwide.

Considering that more than 40 years of information was analyzed, the results showed that on a global scale, the primary data on marine fish available on GBIF and OBIS platforms are still far from being representative and complete. Only 1.14% of the records were useful for our analyses. In addition, we found that the information seems to be biased towards coastal areas, regions close to developed countries and areas where there is a large fishing activity. Finally, the best represented species and families are those with a small body size, which use shallow habitats and have commercial cultural value.

\*Corresponding author

Email address: horacio@ecoinformatica.cl (Horacio Samaniego)

Preprint submitted to Ecological Informatics

February 3, 2023

**Keywords:** Species Richness, Spatial and Temporal Representativeness, Marine Fish, Ecological Information Biases, Ecoinformatics

---

### **3 1. Introduction**

Currently, the more than 243,000 species included in the World Register of Marine Species database ([WORMS, 2022](#)) suggests that only 11 to 78% of all marine species have been discovered revealing a striking picture of vastly incomplete knowledge that may have serious implications for marine conservation ([Luyapaert et al., 2020](#)). Moreover, ongoing climate change represents one of the greatest threats to biodiversity ([Malhi et al., 2020; Turner et al., 2020](#)) and has already been documented to modify the distribution of marine species ([Lenoir et al., 2020](#)). Some of the described effects consider the invasion of non-native species leading to massive species turnover that may lead to the local extinction of large proportions of species ([Cheung et al., 2009](#)).

While species richness is often used to represent diversity patterns, species richness is, in itself, an aggregate variable subsuming the variety of life ([Marquet et al., 2004](#)). Hence, several attempts have focused on the development of more encompassing indices sparking interesting scientific debates to describe ecological heterogeneity ([Tuomisto, 2011; Moreno and Rodríguez, 2011; Daly et al., 2018](#)). However, scientific literature seem to have opted to shift focus to the consequences of biodiversity loss by fostering the usage of new terminology to provide hands-on concepts designed to convey messages to policy makers summarizing the richness of biodiversity ([Pereira et al., 2013; Butchart et al., 2010](#)). Still, while scientists have debated the use of biodiversity terminology, species richness provides a succinct, and easy-to-handle description of the variability across several other quantities describing biodiversity in space and time ([Appeltans et al., 2012](#)), and is, together with other diversity indices, an essential feature to understand how diversity changes under the impact of natural and anthropogenic factors on biomes, regions and ecosystems ([Troia and McManamay, 2017; Magurran and McGill, 2011](#)).

Likewise, biodiversity can also be assessed through life history traits, which are modulated by both evolutionary factors and the variation in habitats and ecosystems ([Neigel, 1997; Hutchings and Baum, 2005](#)). We now know that biodiversity is more likely an expression of the heterogeneity of such life history traits. [Alò et al. 2021](#), for example, shows that while some of the fish diversity is certainly due to environmental processes, a large fraction of such richness variance is also determined by evolved life history traits related, for example, to migratory habits. Therefore, evaluating how life history traits impact richness metrics should deepen our understanding of fish diversity patterns.

While still short of having a robust and standardized biodiversity infrastructure ([Heberling et al., 2021](#)), a great diversity of online repositories with taxonomic information and species occurrences data exist. Among the most important databases hosting marine information are FishBase, a platform that hosts information on the taxonomy of fish, their ecology, trophic information, habitat and history of uses dating back to more than 250 years ([Froese and Pauly, 2000](#)); the Global Biodiversity Information Facility (GBIF), a platform that stores and allows

for the free access to species occurrence records from around the world. GBIF is currently one of the repositories hosting the largest amount of such data in the world (Telenius, 2011; 3 GBIF: The Global Biodiversity Information Facility , 2021); and finally Ocean Biodiversity Information System (OBIS), which houses data on the occurrence and abundance of species from exclusively marine environments (OBIS: Ocean Biodiversity Information System, 2021). 6 Records entered in these repositories are often used for research related to biodiversity assessment, taxonomic reviews, red listing of threatened species, species distribution, and generation 9 of ecological niche models, among others (Yesson et al., 2007). GBIF currently offers more than 1.62 billion occurrence records and OBIS more than 63 million, which increase considerably each year (GBIF: The Global Biodiversity Information Facility , 2021; OBIS: Ocean 12 Biodiversity Information System, 2021).

The records of both platforms come from a wide variety of sources collected following different methodologies at different temporal and spatial scales introducing a great variety biases (Beck et al., 2014; Zizka et al., 2020). Among these, three main types of biases have been 15 described: (i) taxonomic, this occurs when some species and/or families are better sampled than other rarer species (Chandler et al., 2017); (ii) geographic, when data input is unevenly distributed across geographic regions and may prove to obscure inter-region comparisons (Yang 18 et al., 2013; Yesson et al., 2007); and (iii) temporal, which may be prevalent when comparing different time periods as data coverage is unevenly distributed over time (Chandler et al., 2017; Yang et al., 2013). While these biases introduce some uncertainty regarding reliability of 21 species richness descriptions obtained from online platforms (Beck et al., 2014; García-Roselló et al., 2015), they have largely been used to provide an extensive overview of macro-ecological patterns of distribution not available otherwise (Mora et al., 2008; Troia and McManamay, 24 2017).

Still, identifying how sampling effort is distributed across space and time will help to interpret 27 biodiversity patterns and reduce biases. This may be achieved through different weighting schemes for records in areas with sufficient sampling that provide a more reliable contribution compared to underrepresented regions (Phillips et al., 2009; Hortal et al., 2008; Yang et al., 2013).

We here assessed the spatial and temporal representativeness of marine fish records available 30 in the global GBIF and OBIS repositories at the level of marine bioregions in order to pinpoint the location of records that best quantify the diversity of marine fishes. The result is a spatial 33 representativeness analysis that we then overlay on marine conservation areas (UNEP-WCMC and IUCN, 2022) and fisheries exploitation areas (FAO, 2014) to learn whether marine conservation efforts, as well as large fisheries, are located in areas of high species richness or in areas 36 of insufficient data coverage.

Finally, we also analyzed the potential effect that some attributes could have on the incidence 39 of more records in global database repositories. Specifically, we evaluated three hypotheses related to body size, habitat depth and commercial use. The underlying hypotheses are that a better representation in the online platforms may be due to the over sampling of larger

fish, caused by its easy identification; that shallow areas provide easy access to sampling; and economic and commercial interest have elicit a larger representation of culturally relevant species among online biodiversity repositories.

## 2. Methods

### 2.1. Species data

We use all recorded occurrences from the order Actinopterygii hosted in GBIF and OBIS repositories (GBIF.org, 2021; OBIS.org, 2021). Following Alò et al. (2021), evolutionarily older taxa, such as Cephalaspidomorphi, were excluded from this analysis. The libraries *rgbif* and *robis* of the statistical package R were used for data extraction (Chamberlain, 2017; Provoost and Bosch, 2020; R Core Team, 2018). To minimize errors associated with the public usage of GBIF and OBIS repositories, we curated the dataset following Zizka et al. (2020) and filtered the dataset by the columns labeled “scientific name”, “family”, “year”, “Longitude” and “Latitude”. We retained all taxonomic information down to the species level. Any record with NA values was removed. We also removed any duplicated record with identical latitude and longitude as well as any record collected before 1980 (see Alò et al., 2021; García-Roselló et al., 2015). Each record was further assigned to a marine bioregions following Costello and Chaudhary (2017). Spatial data wrangling and plotting was performed with the aid of the following libraries: *sf*, *dplyr*, and *cartography* (Graud and Lambert, 2016; Pebesma, 2018; Wickham et al., 2021). We finally labeled and removed any exotic species record using the *distribution()* function provided by the *rfishbase* library (Boettiger et al., 2012; Froese and Pauly, 2021). To limit our analysis to species occurring within their native range each record was checked against the classification of FAO fisheries area for consistency (FAO, 2014).

### 2.2. Data Analysis by Bioregion

Once the database was cleaned, a subset of the data was created for each of the 30 bioregions. For each bioregion, a count of records, species and families was made, and the Shannon diversity index was calculated using the *vegan* library in R (Oksanen et al., 2020).

#### 2.2.1. Spatial Representativeness Analysis

To assess the spatial representativeness of the data, bioregions were gridded into hexagonal cells of equal area to maximize fit to bioregions areas. These  $1^\circ \times 1^\circ$  hexagonal lattice yielded 60,853 cells in total. We evaluated, in the appendix, two additional spatial resolutions, of  $5^\circ$  and  $10^\circ$  lattice with a total of 3,021 and 958 cells respectively, in order to analyze different biodiversity macropatterns (Tittensor et al., 2010). The expected species richness ( $S_{exp}$ ) was computed as the average between three non-parametric richness estimators so that  $S_{exp} = \frac{1}{3} \sum_i^3 S_i$ , where  $S_i$  is Chao2 ( $S_{chao}$ ), Bootstrap ( $S_{bootstrap}$ ) and Jackknife 1 ( $S_{jackknife1}$ ) (see Magurran and McGill, 2011, for individual definition of indices). Such averaging seeks to minimize biases and potential errors of under- or over-estimation by using a single richness estimator following the work of (Mora et al., 2008; Troia and McManamay, 2017).

We then produced a spatial representativeness index (SRI) by comparing the observed richness ( $S_{obs}$ ) per cell to  $S_{exp}$  (Troia and McManamay, 2017).

$$SRI_i = \frac{S_{obs}}{S_{exp}}, \quad (1)$$

- 3 indicates the degree of representativeness of records to quantify the actual species richness in  
each cell ( $i$ ). Its value ranges from 0 to 1, where 0 represents an unsampled cell and 1 represents  
a fully sampled one (Eq. 1).  
6 Because SRI somehow shows how databases depict the actual species richness, we may  
further classify SRI into four classes labeled by the levels of species richness knowledge they  
represent. Some cell may show very *few* knowledge with  $SRI \in (0, 0.60)$ . Conversely, others  
9 may show to have a *sufficient* species diversity knowledge level for a complete representation  
of species diversity if  $SRI \in (0.60, 0.85)$ . While others, in turn, will have an *adequate* rep-  
resentativeness level if  $SRI \in (0.85, 1.00)$ . Cells with *no records* were also considered as an  
12 independent class.

### 2.2.2. Temporal Representativeness Analysis

We generated plots of cumulative records over time to analyze the temporal distribution of  
15 data records for each bioregions. Accumulation curves for the 30 bioregions were calculated  
based on the observation records and the year of collection. We asses the completeness of the  
sample by evaluating the final 10% tip of the curve using a linear fit after rescaling  $S$  to create  
18 statistically comparable slope units. Slopes close to 0 indicate sufficiently sampled bioregions,  
while slopes closer to one are indicative of insufficient sampling efforts across time.

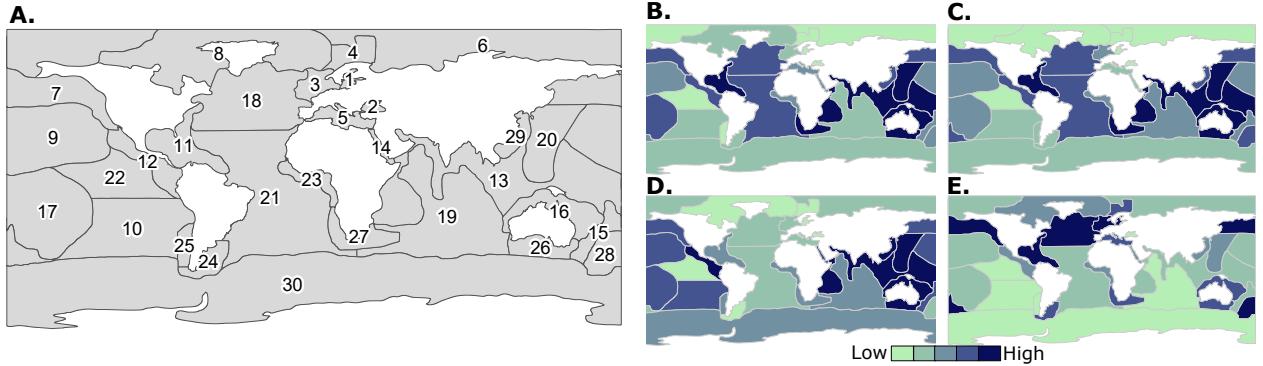
### 2.2.3. Gap Analysis

21 We overlaid the spatial representativeness map (§2.2.1) with shapefiles of Marine Protected  
Areas (MPA) (UNEP-WCMC and IUCN, 2022) and fishing exploitation areas reported by  
(FAO, 2014). The superposition of these layers allowed us to calculate the extent of protection  
24 offered by MPA for each bioregions on a cell basis, and the extent of cells in designated fishing  
zones. This exercise allows to jointly assess the relationship between two opposing human  
impacts and current uncertainties about marine fish diversity.

### 27 2.2.4. Bias Assessment

The evaluation of potential biases generated by body size, habitat depth and cultural value of  
species (§2.1) was assessed from the fishbase database (Froese and Pauly, 2021). Size frequencies  
30 were determined using 80 cm intervals and ranges of habitat depth were determined according  
to the classification of oceanic layers (i.e. epipelagic = 0 - 200 m, mesopelagic = 200 - 1,000 m,  
and bathypelagic= 1,000 - 4,000) (Costello et al., 2010). Parametric correlation analysis was  
33 employed describe the relationship between the frequency of representation, using a logarithmic  
transformation, and the body size and habitat depth, while a simple pie chart shows the  
frequency of cultural values associated to the data.

36 All data and scripts are available (see Appendix A).



**Fig. 1:** Marine bioregions and spatial diversity distribution used in this study. **A.** The 30 marine bioregions from [Costello and Chaudhary \(2017\)](#) used in this study. Number are identification labels in Table 1. **B.** Overall species richness across bioregions; **C.** Family richness; **D.** Average richness (see §2.2.1); and **E.** Shannon diversity index. Note that values in B-E have been normalized for display purposes. See Table 1 for actual values.

### 3. Results

#### 3.1. Records by Bioregions

Just about 1.14% of the total published occurrences in the order Actinopterygii were retained in our analysis. That is, from the 71,670,596 downloaded records off the GBIF and OBIS repositories, 820,004 were considered useful (see Appendix A). This subset consisted of 10,371 species in 361 families. The most represented families in our dataset are Scombridae, Pleuronectidae and Gadidae with 103,762, 57,018 and 52,079 records respectively. The species with the largest representation frequency are *Hippoglossoides platessoides*, *Mola mola* and *Coryphaena hippurus* with 30,885, 21,042 and 21,089 records.

The analysis at bioregion level (Table 1) shows a large variability. The count of records varies across three orders of magnitudes, that is from  $2.68 \times 10^5$  records in the Caribbean Sea and Gulf of Mexico (11) down to  $1.02 \times 10^2$  in the Black Sea (2). The bioregion with the largest species richness and diversity index is the Indo-Pacific Seas and Indian Ocean (13) with  $2.95 \times 10^3$  recorded species and a Shannon index of 6.93 followed by the Coral Sea bioregion (16) with  $2.93 \times 10^3$  species and a Shannon index of 6.75. Likewise, the Coral Sea also presents the largest number of families. It is interesting to note that, while being the largest bioregion (i.e. in  $\text{km}^2$ ), the Southern Ocean show the fewest number of records and the lowest number of species and families across all bioregions. Black Sea (2) and Norwegian (4) are the bioregions with lowest number of record and Shannon index value respectively. Fig. 1 illustrates the location of the 30 marine bioregions and their respective richness and diversity values.

#### 3.2. Geographic Analysis

Fig. 2 shows the cell classification according to SRI (§2.2.1). As expected, no bioregion is completely sampled at the  $1^\circ$  scale resolution. In fact, at this resolution scales, large empty regions with no records are observed. The bioregions with the largest area classified as *Adequate* are the Northeast Atlantic (3) (38.6%), the Caribbean and Gulf of Mexico (11) (22.8%) and the Inland Baltic Sea (1) (21.8%). It should be noted that such cells mainly correspond to coastal areas in the northern hemisphere. On the other hand, the bioregions that present a greater

**Table 1:** Counts of records, species, families and Shannon diversity for each bioregion. The largest values are highlighted in bold.

ID	Bioregion	Area (km <sup>2</sup> )	Records	Species	Families	Shannon
1	Inner Baltic Sea	415,265	8,902	72	30	2.46
2	Black Sea	537,390	102	37	22	3.21
3	NE Atlantic	2,053,367	87,377	310	104	3.90
4	Norwegian Sea	1,132,035	3,046	93	35	2.16
5	Mediterranean	2,859,080	12,532	372	101	3.39
6	Arctic Seas	10,276,384	2,506	114	23	3.90
7	North Pacific	12,974,899	78,070	839	156	4.50
8	North American boreal	8,001,411	9,709	162	48	2.99
9	Mid-tropical N Pacific Ocean	32,685,992	9,310	615	127	4.59
10	South-east Pacific	21,952,036	386	190	89	4.97
11	Caribbean and Gulf of Mexico	8,427,282	<b>268,066</b>	1,703	209	4.49
12	Gulf of California	6,184,523	7,639	885	148	5.93
13	Indo-Pacific seas and Indian Ocean	37,090,022	16,967	<b>2,947</b>	215	<b>6.93</b>
14	Gulfs of Aqaba, Aden, Suez, Red Sea	830,183	926	352	72	5.51
15	Tasman Sea	3,592,495	1,003	380	120	5.36
16	Coral Sea	7,658,020	40,107	2,929	<b>249</b>	6.75
17	Mid South Tropical Pacific	23,418,352	6,083	811	123	5.18
18	Offshore and NW North Atlantic	16,012,294	130,994	897	190	3.46
19	Offshore Indian Ocean	31,076,432	1,263	337	116	4.06
20	Offshore W Pacific	10,291,385	6,363	1,839	232	6.81
21	Offshore S Atlantic	41,435,619	11,960	990	188	3.79
22	Offshore mid-E Pacific	13,815,068	687	79	37	3.04
23	Gulf of Guinea	3,325,856	6,816	384	138	3.95
24	Argentina	2,665,312	8,701	115	52	2.83
25	Chile	1,739,182	250	100	54	4.36
26	Southern Australia	3,824,338	15,643	1,011	201	5.75
27	Southern Africa	4,371,199	19,954	1,142	210	4.16
28	New Zealand	6,293,426	53,879	558	154	3.66
29	North West Pacific	2,457,257	1,767	869	182	6.46
30	Southern Ocean	<b>62,161,808</b>	8,996	294	57	3.98

surface without records correspond to the Southeast Pacific (10) (97.5%), the Southern Ocean (30) (95.7%) and the South Pacific, Mid Tropical (17) (93.9%). Additional results for  $5^\circ \times 5^\circ$  and  $10^\circ \times 10^\circ$  spatial resolution grids are available in Appendix B.

### 3.3. Temporal Analysis

Bioregions show similar trends of data accumulation across the four decades analyzed here (Fig. 3). While a significant increase is apparent in the time period between 2005 and 2010, such increase is not significant for 14 out the 30 bioregions. The Caribbean and Gulf of Mexico (11) is the bioregion with the largest increases in data contribution to the dataset, while the Black Sea (2) is the bioregion with the lowest rate of data contribution in the 40 years span between 1980 and 2020. (See Appendix C for further analysis).

We categorize the slopes of the final 10% of each accumulation curves in Fig. 4. Fourteen bioregions show a slope less than 1. The Mediterranean Sea (5) stands out with the lowest slope value (0.47), while the Black Sea (2) is the bioregion with the steepest final slope (3.13).

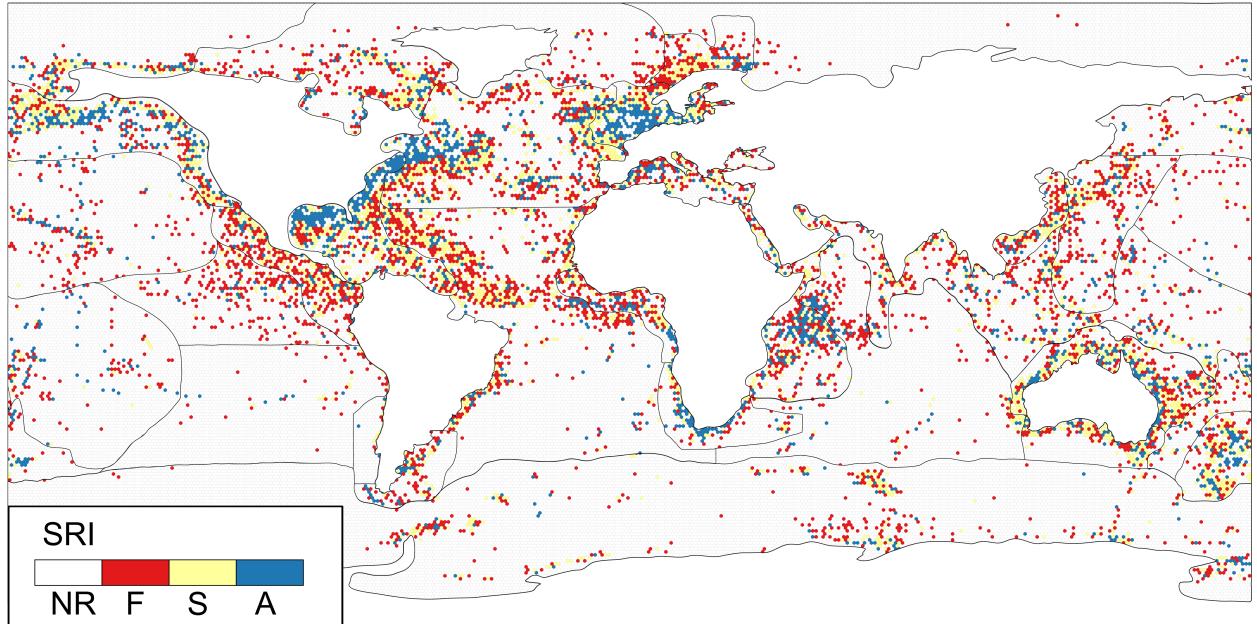
### 3.4. Gap analysis and fishing exploitation areas

Most PAs are concentrated in the North Atlantic bioregions close to the European continent (Table 2). The largest incidences of PAs across bioregions are for the Northeast Atlantic (3) followed by the Inner Baltic Sea (1) ( $n=296$ ) and the Mediterranean Sea (5) ( $n=214$ ) with 734, 296 and 214 PAs respectively. However, the bioregions with the largest area covered by protected areas are the Coral Sea (16), the northeast Atlantic (3) and New Zealand (28) covering a 37.3, 17.4 and 16% of their respective areas. Regarding the sampling level of these bioregions, the Northeast Atlantic (3), the offshore and northwest of the North Atlantic (18), and the Mediterranean Sea (5), are the bioregions with the highest number of PAs and which also have more than 50% of their area sampled as *Adequate*. While Indo-Pacific seas and Indian Ocean bioregion have over 80% of its protected area without any data among the analyzed databases (i.e. *NR*). (See Appendix D)

FAO areas with the largest area categorized as *Adequate* correspond to the northwest Atlantic (20.98%), Northeastern part of the Pacific Ocean (11.78 %), and Western part of the Atlantic Ocean (11.41 %) (Table 3). These FAO areas correspond to regions of the Pacific Ocean (North Pacific, North West Pacific, Mid-tropical N Pacific Ocean and Indo-Pacific seas and Indian Ocean, as well the Gulf of California and Caribbean and Gulf of Mexico). Largest FAO areas with *NR* correspond to the Antarctic part of the Pacific Ocean, the Antarctic part of Atlantic Ocean and Southeastern part of the Atlantic Ocean in the Southern Ocean, Offshore S Atlantic and Southern Africa.

### 3.5. Evaluation of Biases

We evaluated biases for body size, habitat depth, and cultural value for 10,371 marine fish species identified in our database (§3.1).

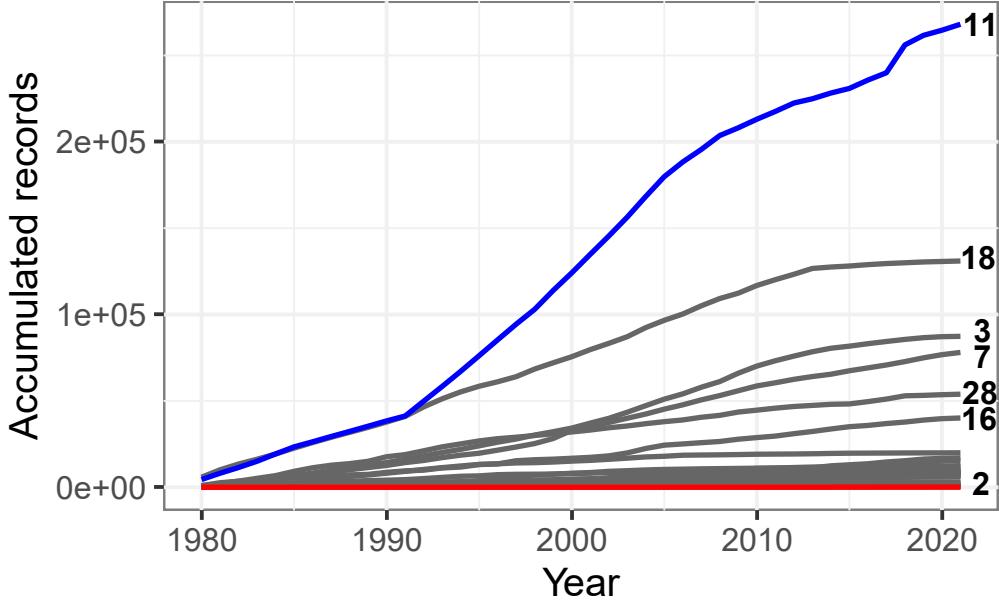


ID	NR	F	S	A	ID	NR	F	S	A
1	18.38	24.96	34.84	21.82	16	43.06	19.17	25.31	12.45
2	79.27	14.56	5.24	0.92	17	93.86	3.22	0.87	2.06
3	24.75	10.19	26.49	38.57	18	59.85	15.45	14.71	10.11
4	45.81	27.66	19.08	7.46	19	93.05	4.25	1.26	1.44
5	51.65	20.66	15.34	12.34	20	72.42	15.53	8.00	4.06
6	92.83	2.67	2.69	1.82	21	79.19	10.95	7.51	2.34
7	70.83	9.34	10.96	8.96	22	82.79	13.04	2.40	1.76
8	77.39	11.79	6.89	3.93	23	47.96	26.19	13.86	11.98
9	92.74	4.80	1.00	1.46	24	71.04	15.37	6.87	6.72
10	97.48	1.62	0.53	0.37	25	93.37	1.96	3.20	1.47
11	35.66	17.20	24.34	22.80	26	63.73	12.85	16.74	6.68
12	43.21	32.32	18.64	5.84	27	74.32	7.34	7.06	11.26
13	75.35	11.51	7.21	5.93	28	66.65	10.24	13.72	9.39
14	49.19	17.62	21.30	11.88	29	46.13	25.20	20.53	8.15
15	81.16	9.81	6.07	2.95	30	95.74	2.24	1.25	0.77

**Fig. 2:** Spatial Representativeness Index (SRI) in  $1^\circ$  hexagonal lattice. Values in the table below indicates the surface area as a percentage of each bioregion for every SRI category (see §2.2.1). **ID** is the identification number given to each bioregion (Table 1). **A** are cells with an adequate representativeness of species richness (i.e.  $\text{SRI} > 0.85$ ). **S** are cells considered as having a sufficient representativeness (i.e.  $\text{SRI} \in (0.60, 0.85)$ ). **F** cells are cells with few records and are thus not considered to be representative of actual species richness (i.e.  $\text{SRI} \in (0, 0.6)$ ). **NR** as cells with no records ( $\text{SRI} = NA$ ).

**Table 2:** Protected area overlap and SRI. “ID” is the identification number given to each bioregion. “PA num” is the number of protected areas (PA). “P”, the percentage of the surface area covered by PA. “A>50%” is the number of protected areas with more than 50% of its surface area classified as *Adequate*. “NR> 80%” is the number of PAs with a surface area larger than 80% with *No Records*. The highest values are highlighted in bold face.

ID	PA num	% area	A >50%	NR>80%
1	296	12.89	92	73
2	10	1.23	2	1
3	<b>734</b>	17.36	<b>614</b>	5
4	7	5.01	0	1
5	214	1.66	104	12
6	2	0.20	0	1
7	32	0.16	27	1
8	7	0.15	1	3
9	27	3.79	14	2
10	2	0.69	0	1
11	97	0.45	47	7
12	7	0.28	0	0
13	165	2.56	63	<b>29</b>
14	2	0.21	1	0
15	16	13.83	0	1
16	86	<b>37.29</b>	22	3
17	13	4.06	4	3
18	197	4.91	131	6
19	8	0.83	0	3
20	122	0.09	29	2
21	40	0.07	1	2
22	1	0.00	1	0
23	2	0.03	0	0
24	8	0.28	4	1
25	3	0.00	1	1
26	125	9.74	19	0
27	45	1.45	39	1
28	76	16.00	39	10
29	19	0.25	1	1
30	11	2.78	0	7



**Fig. 3:** Records accumulation rate for each bioregion across the four decades analyzed. The blue line is the accumulation of fish records in the Caribbean and Gulf of Mexico bioregion (11) and the red line shows the accumulation rate in the Black Sea (2). Numbers as the end of each timeseries correspond to the bioregion ID in Table 1.

### 3.5.1. Body size

The range 0-80 cm is the most frequently occurring size length. Three species stand out with

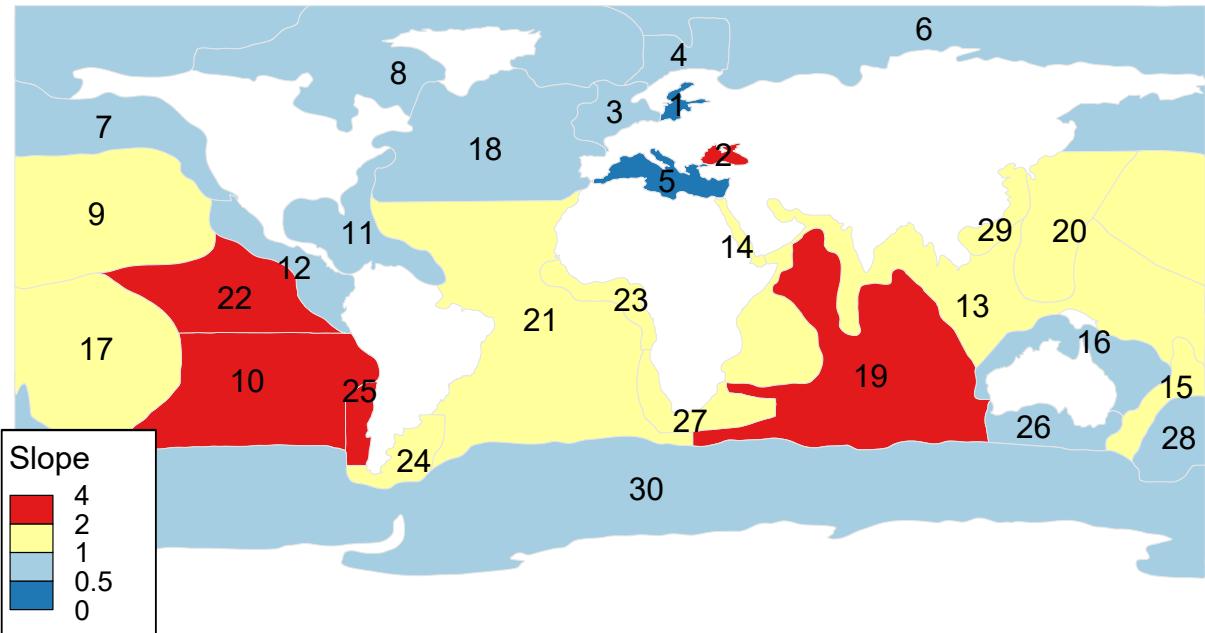
- 3 the highest numbers of records, *Scomber scombrus*, *Lagodon rhomboides* and *Mallotus villosus* with 20,995, 19,563 and 13,609 records respectively. These species are distributed mainly in the Northeast Atlantic (3) and Offshore and Northwest North Atlantic (18) bioregions. While the
- 6 families that accumulate the greatest number of records correspond to Sparidae , Scombridae and Labridae with 24,837, 21,719 and 21,035 records. These families are mainly distributed in the Caribbean Sea and Gulf of Mexico and the Northeast Atlantic.

### 9 3.5.2. Habitat depth

The most frequent depth range is between 0-838 m (i.e. epipelagic and mesopelagic zones), and the species with the highest number of records are *Mola mola*, *Coryphaena hippurus* and

- 12 *Lagodon rhomboides* with 21,089, 21,042 and 19,563 occurrences in the databases. These species are distributed mainly around the Caribbean Sea and Gulf of Mexico (11) bioregions, as well as the following bioregions: Offshore and NW North Atlantic (18) and the South Atlantic Coast (21).
- 15 The families that accumulate a greater number of records correspond to Scombridae, Gadidae, Sparidae with 63,572, 38,876 and 30,041 records. These are mostly distributed in the northern hemisphere. That is, the Caribbean and the Gulf of Mexico (11), Offshore and NW of the North Atlantic (18) and part of the South Atlantic Ocean Coast (21) bioregions.

Figure 5 shows that body size and habitat depth have a negative correlation with the frequency of records.



**Fig. 4:** Graphical representation of the slope values of the species accumulation curve for each bioregion. The slope corresponds to the final 10% of the species accumulation curve. See §2.2.2 for details regarding the analysis.

### 3.5.3. Cultural value

Finally, when analyzing the most frequent cultural value represented across our dataset

(Fig. 6), “Commercial” use of the species emerges as the most important with a 73.4% among records, followed by the category “No interest” (5.03%), and “Subsistence fishing” (3.08%).

## 4. Discussion

Our work provides a methodological framework based on a set of non-parametric estimators to quantify the potential number of species from incidence data (Chao et al., 2009). We used hexagonal grids that fit the geographic reality of marine ecosystems, and we placed special emphasis on cleaning the occurrence data in their taxonomy (Jin and Yang, 2020) and any potential input errors associated with large and massive datasets (Zizka et al., 2020). This led us to focus on only evaluating marine species in the order Actinopterygii (Alò et al., 2021).

Publicly accessible occurrence records are growing rapidly, partly due to the significant advances in ecoinformatics (Lenoir et al., 2020; Oliver et al., 2021). These databases harbor a growing variety of sources, including museum specimens, field observations, acoustic and visual sensors, and citizen science efforts (Amano et al., 2016). However, despite the incredible accumulation of biodiversity records, not all the data is really useful, nor does it represent new insights into the distribution of species (Bayraktarov et al., 2019; Zizka et al., 2020). That is why a systematic evaluation of the integrity and coverage of this information is required (Troia and McManamay, 2017).

There is an extensive bibliography that evaluates the record quality available for different taxonomic groups. Some examples are: legumes on a global scale (Yesson et al., 2007), lep-

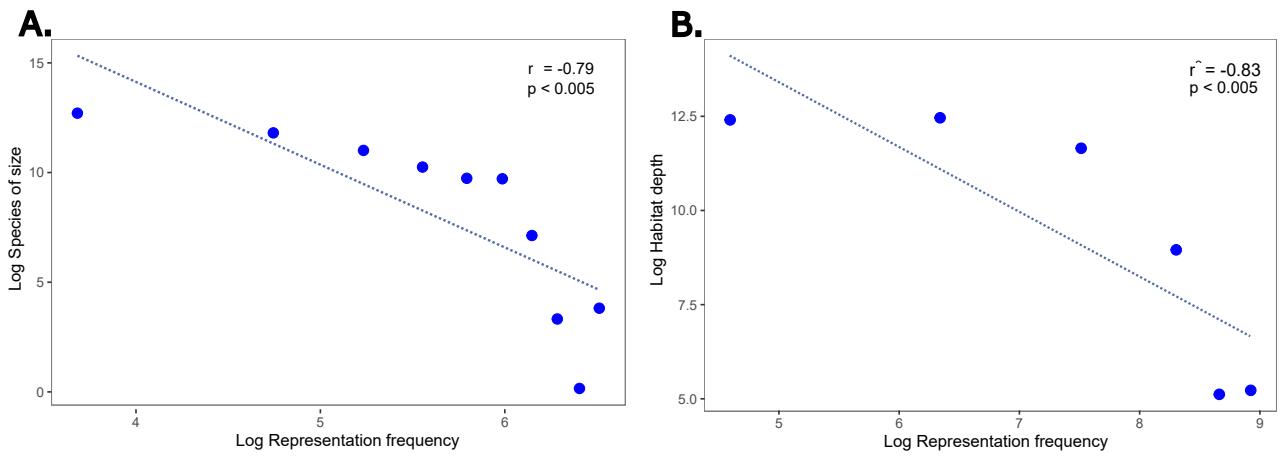
**Table 3:** Results of overlapping FAO fishery exploitation areas and SRI grid. The surface area corresponding to each bioregion, and the percentage of surface area of each classification. Area is in thousands of km<sup>2</sup>; NR is the percentage of cells with *No Records*; F is the percentage of classified cells with *Few* records; S, the percentage of classified cells with *Sufficient* records, and A, the percentage of classified cells with *Adequate* records. The highest values are highlighted in bold.

FAO Area Name	Area	NR	F	S	A
Arctic Ocean	9.24	89.81	4.74	3.29	2.16
Northwestern part of the Atlantic Ocean	6.28	42.48	16.91	19.62	<b>20.98</b>
Northeastern part of the Atlantic Ocean	14.35	62.85	13.42	14.02	9.71
Western part of the Atlantic Ocean	14.65	43.56	<b>22.37</b>	<b>22.66</b>	11.41
Eastern Central part of the Atlantic Ocean	14.19	64.17	20.43	10.07	5.34
Mediterranean Sea and the Black Sea	2.99	57.82	19.07	12.75	10.36
Southwestern part of the Atlantic Ocean	17.70	85.60	7.08	5.42	1.91
Southeastern part of the Atlantic Ocean	19.07	93.34	2.13	1.61	2.93
Antarctic part of the Atlantic Ocean	12.48	95.31	2.38	1.63	0.68
Western part of the Indian Ocean	29.42	79.15	10.42	4.87	5.56
Eastern part of the Indian Ocean	31.21	87.57	5.24	5.16	2.04
Antarctic and Southern of the Indian Ocean	12.75	90.02	5.23	3.13	1.62
Northwestern part of the Pacific Ocean	21.54	80.43	10.87	6.02	2.67
Northeastern part of the Pacific Ocean	7.63	68.73	8.53	10.96	11.78
Northeastern part of the Pacific Ocean	33.36	78.69	9.77	6.88	4.66
Eastern Central part of the Pacific Ocean	<b>48.25</b>	85.09	9.01	3.47	2.43
Southwestern part of the Pacific Ocean	27.64	87.72	4.79	4.88	2.61
Southeastern part of the Pacific Ocean	30.80	92.85	4.89	1.33	0.92
Antarctic part of the Pacific Ocean	10.23	<b>95.61</b>	2.28	0.71	1.40

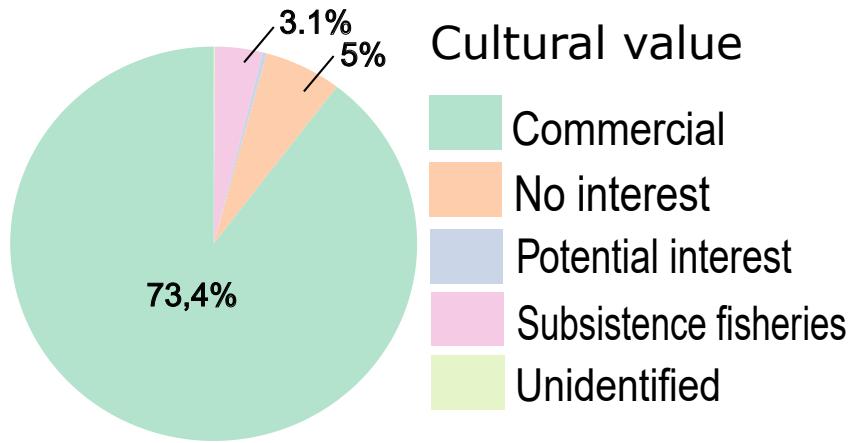
idoptera from Great Britain and woody plants in Panama (Chao et al., 2009), global marine biodiversity (Tittensor et al., 2010), vascular plants in China (Yang et al., 2013), marine fish on

- 3 a global scale (Mora et al., 2008; García-Roselló et al., 2015), freshwater fish in the USA (Troia and McManamay, 2017; Pelayo-Villamil et al., 2018), and terrestrial mammals on a global scale (Oliver et al., 2021), among many others. Assuming that that not all data available in these
- 6 repositories are useful for biodiversity analyses, several efforts have proposed parametric and non-parametric estimators for data cleaning and species richness analysis, ModestR (García-Roselló et al., 2013), KnowBr (Lobo et al., 2018), and RWizard (Guisande and Lobo, 2019)
- 9 among these.

Regarding the units of analysis, here we estimate the species richness at the grid level in order to obtain more uniform results on the distribution of occurrences and avoid overestimating the SRI for marine bioregions (Pelayo-Villamil et al., 2018). In addition, we evaluated two additional grid sizes ( $5^\circ \times 5^\circ$  and  $10^\circ \times 10^\circ$ ), and like other studies, our results show that the coarser the resolution used, the greater the overestimation of the index than richness. , and, conversely, the finer the analysis scale, the more deficient the sampling (Tittensor et al., 2010; García-Roselló et al., 2015; Meyer et al., 2015; Troia and McManamay, 2016, 2017). In view of the above, we recommend using grids that really allow observing macro-ecological patterns, especially in coastal regions, which may be underrepresented when using thick and square grids (Pelayo-Villamil et al., 2018).



**Fig. 5:** A. Relationship between marine fish representation frequency ( $\log_{10} a$ ) in GBIF and OBIS and habitat depth. B. Relationship between the frequency of representation of marine fishes ( $\log_{10} a$ ) in GBIF and OBIS and the size of the species. The dotted line is only there to highlight the negative trend between variables.



**Fig. 6:** Frequency of marine fish representation in GBIF and OBIS according to importance of cultural use.

Considering that more than 40 years of data were analysed, our results demonstrated that on a global scale, the primary marine fish data available on the GBIG and OBIS platforms are still far from being representative and complete. Compared with other studies evaluating the same taxonomic group (Mora et al., 2008; García-Roselló et al., 2015), although we obtained similar macroecological patterns, only 1.14% of the records extracted from both repositories were useful for our analyses. The large percentage of the occurrences presented input errors or did not have the necessary data to generate a reliable analysis (Yesson et al., 2007; García-Roselló et al., 2014).

We also found evidence of strong information biases in the records explored. On the one hand, when analyzing the families and species with the greatest representation, they coincide with groups of fish of commercial interest, demonstrating the existence of **taxonomic bias** of the data (Melo-Merino et al., 2020). This is the case of the families Scombridae, Pleuronectidae and Gadidae, which include species of nutritional importance such as tuna, cod, haddock, among others (Cohen et al., 1990). The same is true for the species with the largest number of records, *H. platessoides* (Pleuronectidae), *C. hippurus* (Coryphaenidae), and *M. mola* (Molidae), the first two are species exploited by the fishing industry, with the exception of sunfish (*M. mola*) which has a wide distribution and is mostly associated with scientific and

recreational interest (Pope et al., 2010).

The unequal contribution of data at the spatial level is another factor that must be considered to work with data available on ecoinformatic platforms. There is a clear preference for certain regions and/or ecosystems as a result of **geographical bias**. The literature indicates that the highest data contribution rates correspond to developed countries (Yesson et al., 2007; Chandler et al., 2017), and those coastal regions with better road connectivity (Chandler et al., 2017; Melo-Merino et al., 2020). This information uncertainty is also particularly prevalent in under-sampled marine habitats, such as the deep sea (Webb et al., 2010). Our results coincide with what is described in the literature, regardless of the size of the grid that was used to generate the analysis, the bioregions that include the Atlantic, the Caribbean and the Gulf of Mexico, and the Baltic Sea are the regions with the highest number of area sampled as *Adequate* associated mainly with coastal areas. While the bioregions that include the South and Southeast Pacific (including the southern coast of South America), the Southern Ocean, and the Arctic Seas are the regions with the least spatial representativeness of records, the proportion of cells without records (*NR*) exceeds 90%. This large area without samples will make any attempt to describe species richness and distribution highly unreliable in these bioegions (Yang et al., 2013; Troia and McManamay, 2017). The marine regions that include the water column, the seabed and the subsoil beyond the limits of jurisdiction of countries cover almost half of the Earth's surface and support a great abundance and diversity of life (Visalli et al., 2020). Even so, when examining the marine ichthyofauna occurrence data, these represent the least sampled areas.

Finally, the **time bias** of the data is also present in our study. Diametric differences in species identification and sampling methodologies over the decades have resulted in the production of databases of variable quality. However, the current era is characterized by more accurate data thanks to improvements in individual capture and identification tools (Costello et al., 2015; Jin and Yang, 2020). For these reasons our approach considers occurrence records since 1980, however, the coverage of occurrence data is uneven over time when comparing between marine bioregions. Despite evaluating more than four decades of data, still 46% of marine bioregions have insufficient sampling efforts. Not surprisingly, the Caribbean and Gulf of Mexico (11) bioregion is the region with the largest input of data, demonstrating once again that geographic sampling bias has strong effects on spatial predictions of species richness (Yang et al., 2013). Future sampling efforts should focus on those bioregions corresponding to low or equatorial latitudes, areas where biogeographic studies show that marine biodiversity is concentrated (Costello and Chaudhary, 2017).

All the biases that we have described, added to the inherent problems in data capture, foster and deepen various information gaps that affect the effective spatio-temporal quantification of biodiversity (Magurran and McGill, 2011). In this study, we have overlapped our estimates of species richness with the global marine protected areas declared up to the beginning of the year 2022 (UNEP-WCMC and IUCN, 2022), and the areas of fishing exploitation reported by the FAO (FAO, 2014). This exercise demonstrates how important it is to have public

databases that can faithfully reflect the taxonomic and biogeographical knowledge available for each region of the world ([Pelayo-Villamil et al., 2018](#)). Our results indicate that Northeastern 3 Atlantic (3) bioregion has the largest number of protected areas (n=734), of which the vast majority (n=614) have over 50% of their area *Adequate* sampled. However, this bioregion is not considered a marine biodiversity hotspot, compared to the Coral Sea and New Zealand 6 bioregions, which have a larger area protected by marine conservation areas ([Ramírez et al., 2017](#)). However, we found a low proportion of well-sampled cells in both regions, demonstrating 9 the existence of important information gaps, at least for fish of the order Actinopterygii. We emphasize the need to correct these information gaps so that conservation efforts that seek the implementation of new marine protection areas can have reliable data so as not to underestimate 12 the biodiversity of species ([Sala et al., 2021](#)).

In the same way, by overlapping the bioregions with the fishing exploitation zones, we 15 determined that the North Pacific (7) North West Pacific (29), Mid-tropical N Pacific Ocean (9) and Indo-Pacific seas and Indian Ocean (13) bioregions , as well the Gulf of California (21) and Caribbean and Gulf of Mexico (11) are the regions with the highest representation of the 18 data and where fishing activity is concentrated. According to ([Kroodsma et al., 2018](#)), the area corresponding to the central Atlantic and Northeast Pacific present little intense fishing effort, while the regions associated with the Northeast Atlantic, the Northeast Atlantic (Europe) 21 regions, and the Northwest Pacific are known to have a huge fishing development and that is where fishing efforts are concentrated worldwide. The southeastern Atlantic Ocean (FAO area 47 and 88), part of the Pacific Ocean (FAO area 88) and Antarctica (FAO area 48 and 88) are the regions with the highest percentage of cells without records ( $NR = >93\%$ ). When compared 24 with the findings of ([Kroodsma et al., 2018](#)), these areas agree with the “holes” without fishing effort data, which is explained by the geographical remoteness and the lack of technological development necessary for the fisheries to extend to new domains ([Visalli et al., 2020](#)). This limits both the exploitation of marine resources and the collection of data.

The hypotheses that we evaluated in this work were necessary to understand what the data 27 collection trends have been and to be able to take future actions to correct the biases described. Our first hypothesis about the size of the body of the fish was rejected. Small fish species (0-80 cm) are the ones that accumulate the largest number of records, among which *S. scombrus* 30 (Scombridae), *L. rhomboides* (Sparidae), *M. villosus* (Osmeridae), are <50 cm species that stand out for presenting the largest number of records, and, in addition, they are distributed 33 in the best sampled regions (Mediterranean Sea, Gulf of Mexico and the Caribbean, and the Atlantic Ocean). The size of the fish is inversely proportional to the abundance and, therefore, to the frequency of human use, both scientific and commercial ([Pauly and Palomares, 2005](#)). This 36 difference in the sampling effort generates an evident overrepresentation of the smaller species and therefore deepens the taxonomic bias. The hypothesis about the depth of the habitat is accepted, at less depth there is a greater representation of species of marine fish. The 39 pelagic zone has a high concentration of data and effectively corresponds to shallow regions and therefore easily accessible, which generates all the conditions for data collection ([Melo-Merino](#)

et al., 2020). It has been pointed out that the concentration of species decreases as the depth of the ocean increases, however, it is precisely these areas that have been least sampled and where 3 there is the greatest probability of discovering new species (Costello and Chaudhary, 2017). This demonstrates the need to concentrate efforts on the deeper regions of the water column (mesopelagic, bathyal, and abyssal) for a more equitable representation of marine ecosystems. 6 Finally, the hypothesis of the use of the species is also accepted. The species of marine fish that have a more beneficial or lucrative use for humans are better represented in the analyzed databases. We believe that this is related to the fact that the fishing industry is one of the 9 main sources of information for platforms such as OBIS (Zhang and Grassle, 2002).

Today, marine ecosystems and their biodiversity face the great challenges of climate change and the impact of human activity, especially those species considered key food resources for 12 survival (Hollowed et al., 2013; Ramírez et al., 2017; O'Hara et al., 2021). It is necessary to focus and strengthen the study of those areas with very few or no records, since the descriptions of the geographic ranges of the species and their temporal dynamics are fundamental measures 15 for the evaluation of the real state of biodiversity (Lenoir et al., 2020; Oliver et al., 2021). Having more reliable data will allow effective conservation actions to be implemented.

#### *Acnlowledgement*

18 Funding for this research was provided by the National Agency of Research and Development of Chile through project FONDECYT Regular #11211490 to HS. We thank professor Ricardo Giesecke for valuable comments on earlier versions of this manuscript.

#### **21 References**

- Alò, D., Lacy, S.N., Castillo, A., Samaniego, H.A., Marquet, P.A., 2021. The macroecology of fish migration. *Global Ecology and Biogeography* 30, 99–116. doi:[10.1111/geb.13199](https://doi.org/10.1111/geb.13199).
- 24 Amano, T., Lamming, J.D., Sutherland, W.J., 2016. Spatial gaps in global biodiversity information and the role of citizen science. *Bioscience* 66, 393–400. doi:[10.1093/biosci/biw022](https://doi.org/10.1093/biosci/biw022).
- Appeltans, W., Ahyong, S., Anderson, G., Angel, M., Artois, T., Bailly, N., Bamber, R., 27 Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C., Brandão, S., Bray, R., Bruce, N., Cairns, S., Chan, T.Y., Cheng, L., Collins, A., Cribb, T., Curini-Galletti, M., Dahdouh-Guebas, F., Davie, P., Dawson, M., De Clerck, O., Decock, W., 30 De Grave, S., de Voogd, N., Domning, D., Emig, C., Erséus, C., Eschmeyer, W., Fauchald, K., Fautin, D., Feist, S., Fransen, C., Furuya, H., Garcia-Alvarez, O., Gerken, S., Gibson, D., Gittenberger, A., Gofas, S., Gómez-Daglio, L., Gordon, D., Guiry, M., Hernandez, F., 33 Hoeksema, B., Hopcroft, R., Jaume, D., Kirk, P., Koedam, N., Koenemann, S., Kolb, J., Kristensen, R., Kroh, A., Lambert, G., Lazarus, D., Lemaitre, R., Longshaw, M., Lowry, J., Macpherson, E., Madin, L., Mah, C., Mapstone, G., McLaughlin, P., Mees, J., Meland, K., 36 Messing, C., Mills, C., Molodtsova, T., Mooi, R., Neuhaus, B., Ng, P., Nielsen, C., Norenburg, J., Opresko, D., Osawa, M., Paulay, G., Perrin, W., Pilger, J., Poore, G., Pugh,

- P., Read, G., Reimer, J., Rius, M., Rocha, R., Saiz-Salinas, J., Scarabino, V., Schierwater, B., Schmidt-Rhaesa, A., Schnabel, K., Schotte, M., Schuchert, P., Schwabe, E., Segers, H., Self-  
3 Sullivan, C., Shenkar, N., Siegel, V., Sterrer, W., Stöhr, S., Swalla, B., Tasker, M., Thuesen, E., Timm, T., Todaro, M., Turon, X., Tyler, S., Uetz, P., van der Land, J., Vanhoorne, B., van Ofwegen, L., van Soest, R., Vanaverbeke, J., Walker-Smith, G., Walter, T., Warren,  
6 A., Williams, G., Wilson, S., Costello, M., 2012. The magnitude of global marine species diversity. *Current Biology* 22, 2189–2202. doi:[10.1016/j.cub.2012.09.036](https://doi.org/10.1016/j.cub.2012.09.036).
- Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E.L., Nguyen, H.A., McRae, L., Possingham, H.P., Lindenmayer, D.B., 2019. Do big unstructured biodiversity data mean more knowledge? *Frontiers in Ecology and Evolution* , 239. doi:[10.3389/fevo.2018.00239](https://doi.org/10.3389/fevo.2018.00239).
- Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the gbif database and  
12 its effect on modeling species' geographic distributions. *Ecological Informatics* 19, 10–15. doi:[10.1016/j.ecoinf.2013.11.002](https://doi.org/10.1016/j.ecoinf.2013.11.002).
- Boettiger, C., Lang, D.T., Wainwright, P., 2012. Rfishbase: exploring, manipulating and  
15 visualizing fishbase data from r. *Journal of Fish Biology* 81, 2030–2039. doi:[10.1111/j.1095-8649.2012.03464.x](https://doi.org/10.1111/j.1095-8649.2012.03464.x).
- Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C., Watson, R., 2010.  
21 Global biodiversity: Indicators of recent declines. *Science* 328, 1164–1168. doi:[10.1126/science.1187512](https://doi.org/10.1126/science.1187512).
- Chamberlain, S., 2017. rgbif: Interface to the global "biodiversity" information facility "api".  
27 r package version 0.9.8. URL: <https://CRAN.R-project.org/package=rgbif>.
- Chandler, M., See, L., Copas, K., Bonde, A.M., López, B.C., Danielsen, F., Legind, J.K.,  
30 Masinde, S., Miller-Rushing, A.J., Newman, G., et al., 2017. Contribution of citizen science towards international biodiversity monitoring. *Biological conservation* 213, 280–294. doi:[10.1016/j.biocon.2016.09.004](https://doi.org/10.1016/j.biocon.2016.09.004).
- Chao, A., Colwell, R.K., Lin, C.W., Gotelli, N.J., 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90, 1125–1133. doi:[10.1890/07-2147.1](https://doi.org/10.1890/07-2147.1).
- Cheung, W.W., Lam, V.W., Sarmiento, J.L., Kearney, K., Watson, R., Pauly, D., 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish and fisheries* 10, 235–251. doi:[10.1111/j.1467-2979.2008.00315.x](https://doi.org/10.1111/j.1467-2979.2008.00315.x).

- Cohen, D.M., Inada, T., Iwamoto, T., Scialabba, N., 1990. Gadiform fishes of the world. FAO Fisheries Synopsis 10, I.
- <sup>3</sup> Costello, M.J., Chaudhary, C., 2017. Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Current Biology* 27, R511–R527. doi:[10.1016/j.cub.2017.04.060](https://doi.org/10.1016/j.cub.2017.04.060).
- <sup>6</sup> Costello, M.J., Cheung, A., De Hauwere, N., 2010. Surface area and the seabed area, volume, depth, slope, and topographic variation for the world's seas, oceans, and countries. *Environmental Science & Technology* 44, 8821–8828. doi:[10.1021/es1012752](https://doi.org/10.1021/es1012752).
- <sup>9</sup> Costello, M.J., Vanhoorne, B., Appeltans, W., 2015. Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology* 29, 1094–1099. doi:[10.1111/cobi.12496](https://doi.org/10.1111/cobi.12496).
- <sup>12</sup> Daly, A.J., Baetens, J.M., De Baets, B., 2018. Ecological diversity: Measuring the unmeasurable. *Mathematics* 6, 119. doi:[10.3390/math6070119](https://doi.org/10.3390/math6070119).
- FAO, 2014. Fao statistical areas for fishery purposes. fao fisheries and aquaculture department [online] URL: <http://www.fao.org/fishery/area/search/en>.
- <sup>15</sup> Froese, R., Pauly, D., 2000. FishBase 2000: concepts designs and data sources. volume 1594. [The WorldFish Center]. URL: <http://hdl.handle.net/20.500.12348/2428>.
- Froese, R., Pauly, D.E., 2021. Fishbase. URL: <https://www.fishbase.org>.
- <sup>18</sup> García-Roselló, E., Guisande, C., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., Vaamonde, A., Granado-Lorencio, C., 2013. Modestr: a software tool for managing and analyzing species distribution map databases. *Ecography* 36, 1202–1207. doi:[10.1111/j.1600-0587.2013.00374.x](https://doi.org/10.1111/j.1600-0587.2013.00374.x).
- <sup>21</sup> García-Roselló, E., Guisande, C., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., González Vilas, L., González-Dacosta, J., Vaamonde, A., Granado-Lorencio, C., 2014. Using modestr to download, import and clean species distribution records. *Methods in ecology and evolution* 5, 708–713. doi:[10.1111/2041-210X.12209](https://doi.org/10.1111/2041-210X.12209).
- <sup>24</sup> García-Roselló, E., Guisande, C., Manjarrés-Hernández, A., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., González-Vilas, L., Vari, R.P., Vaamonde, A., Granado-Lorencio, C., et al., 2015. Can we derive macroecological patterns from primary global biodiversity information facility data? *Global Ecology and Biogeography* 24, 335–347. doi:[10.1111/geb.12260](https://doi.org/10.1111/geb.12260).
- <sup>27</sup> GBIF: The Global Biodiversity Information Facility , 2021. What is gbif? URL: <https://www.gbif.org/what-is-gbif>.
- <sup>30</sup> GBIF.org, 2021. Occurrence download. URL: <https://www.gbif.org/occurrence/download/0039590-210914110416597>, doi:[10.15468/DL.V2PFS3](https://doi.org/10.15468/DL.V2PFS3). last accessed 29 October 2021.

- Giraud, T., Lambert, N., 2016. cartography: Create and integrate maps in your r workflow. *Journal of Open Source Software* 1, 54. doi:[10.21105/joss.00054](https://doi.org/10.21105/joss.00054).
- <sup>3</sup> Guisande, C., Lobo, J., 2019. Discriminating well surveyed spatial units from exhaustive biodiversity databases. r package version. 2.0. URL: <http://cran.r-project.org/web/packages/KnowBR>.
- <sup>6</sup> Heberling, J.M., Miller, J.T., Noesgaard, D., Weingart, S.B., Schigel, D., 2021. Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences* 118, e2018093118. doi:[10.1073/pnas.2018093118](https://doi.org/10.1073/pnas.2018093118).
- <sup>9</sup> Hollowed, A.B., Barange, M., Beamish, R.J., Brander, K., Cochrane, K., Drinkwater, K., Foreman, M.G., Hare, J.A., Holt, J., Ito, S.i., et al., 2013. Projected impacts of climate change on marine fish and fisheries. *ICES Journal of Marine Science* 70, 1023–1037. doi:[10.1093/icesjms/fst081](https://doi.org/10.1093/icesjms/fst081).
- <sup>12</sup> Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A., 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117, 847–858. doi:[10.1111/j.0030-1299.2008.16434.x](https://doi.org/10.1111/j.0030-1299.2008.16434.x).
- <sup>15</sup> Hutchings, J.A., Baum, J.K., 2005. Measuring marine fish biodiversity: temporal changes in abundance, life history and demography. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 315–338. doi:[10.1098/rstb.2004.1586](https://doi.org/10.1098/rstb.2004.1586).
- <sup>18</sup> Jin, J., Yang, J., 2020. Bdcleaner: A workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. *Global Ecology and Conservation* 21, e00852. doi:[10.1016/j.gecco.2019.e00852](https://doi.org/10.1016/j.gecco.2019.e00852).
- <sup>21</sup> Kroodsma, D.A., Mayorga, J., Hochberg, T., Miller, N.A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T.D., Block, B.A., et al., 2018. Tracking the global footprint of fisheries. *Science* 359, 904–908. doi:[10.1126/science.aao5646](https://doi.org/10.1126/science.aao5646).
- <sup>24</sup> Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., Grenouillet, G., 2020. Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution* 4, 1044–1059. doi:[10.1038/s41559-020-1198-2](https://doi.org/10.1038/s41559-020-1198-2).
- <sup>27</sup> Lobo, J.M., Hortal, J., Yela, J.L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., Guisande, C., 2018. Knowbr: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators* 91, 241–248. doi:[10.1016/j.ecolind.2018.03.077](https://doi.org/10.1016/j.ecolind.2018.03.077).
- <sup>30</sup> Luypaert, T., Hagan, J.G., McCarthy, M.L., Poti, M., 2020. Status of marine biodiversity in the anthropocene, in: YOUMARES 9-The Oceans: Our research, our future. Springer, Cham, pp. 57–82. doi:[10.1007/978-3-030-20389-4\\_4](https://doi.org/10.1007/978-3-030-20389-4_4).

- Magurran, A.E., McGill, B.J., 2011. Biological diversity: frontiers in measurement and assessment. Oxford University Press. doi:[10.1086/666756](https://doi.org/10.1086/666756).
- <sup>3</sup> Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M.G., Field, C.B., Knowlton, N., 2020. Climate change and ecosystems: Threats, opportunities and solutions. doi:[10.1098/rstb.2019.0104](https://doi.org/10.1098/rstb.2019.0104).
- <sup>6</sup> Marquet, P.A., Fernández, M., Navarrete, S.A., Valdovinos, C., 2004. Diversity emerging: towards a deconstruction of biodiversity patterns, in: Lomolino, M., Heaney, L. (Eds.), Frontiers of Biogeography: New directions in the Geography of Nature. Cambridge University Press, pp. 191–209.
- <sup>9</sup> 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. doi:[10.1016/j.ecolmodel.2019.108837](https://doi.org/10.1016/j.ecolmodel.2019.108837).
- Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. *Nature communications* 6, 1–8. doi:[10.1038/ncomms9221](https://doi.org/10.1038/ncomms9221).
- <sup>15</sup> Mora, C., Tittensor, D.P., Myers, R.A., 2008. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society B: Biological Sciences* 275, 149–155. doi:[10.1098/rspb.2007.1315](https://doi.org/10.1098/rspb.2007.1315).
- <sup>18</sup> Moreno, C.E., Rodríguez, P., 2011. Do we have a consistent terminology for species diversity? back to basics and toward a unifying framework. *Oecologia* 167, 889–892. doi:[10.1007/s00442-011-2125-7](https://doi.org/10.1007/s00442-011-2125-7).
- <sup>21</sup> Neigel, J., 1997. Marine Biodiversity: Patterns and Processes. Cambridge, Cambridge University Press. chapter Population genetics and demography of marine species. URL: <http://www.cambridge.org/9780521552226>.
- <sup>24</sup> OBIS: Ocean Biodiversity Information System, 2021. About obis URL: <https://obis.org/>. OBIS.org, 2021. Occurrence download. URL: <https://datasets.obis.org/downloads/9fd73b2a-cf6f-4ef9-a0e3-2d1f653520d3.zip>. last accessed 29 October 2021.
- <sup>27</sup> Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2020. The vegan package URL: <https://github.com/vegandevels/vegan>.
- <sup>30</sup> Oliver, R.Y., Meyer, C., Ranipeta, A., Winner, K., Jetz, W., 2021. Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. *PLoS Biology* 19, e3001336. doi:[10.1371/journal.pbio.3001336](https://doi.org/10.1371/journal.pbio.3001336).
- <sup>33</sup> O'Hara, C.C., Frazier, M., Halpern, B.S., 2021. At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts. *Science* 372, 84–87. doi:[10.1126/science.abe6731](https://doi.org/10.1126/science.abe6731).

- Pauly, D., Palomares, M.L., 2005. Fishing down marine food web: it is far more pervasive than we thought. *Bulletin of marine science* 76, 197–212.
- <sup>3</sup> Pebesma, E.J., 2018. Simple features for r: standardized support for spatial vector data. *R J.* 10, 439. doi:[10.32614/RJ-2018-009](https://doi.org/10.32614/RJ-2018-009).
- <sup>6</sup> Pelayo-Villamil, P., Guisande, C., Manjarrés-Hernández, A., Jiménez, L.F., Granado-Lorencio, C., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., Lobo, J.M., 2018. Completeness of national freshwater fish species inventories around the world. *Biodiversity and Conservation* 27, 3807–3817. doi:[10.1007/s10531-018-1630-y](https://doi.org/10.1007/s10531-018-1630-y).
- <sup>9</sup> Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science* 339, 277–278. doi:[10.1126/science.1229931](https://doi.org/10.1126/science.1229931).
- <sup>15</sup> Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications* 19, 181–197. doi:[10.1890/07-2153.1](https://doi.org/10.1890/07-2153.1).
- <sup>18</sup> Pope, E.C., Hays, G.C., Thys, T.M., Doyle, T.K., Sims, D.W., Queiroz, N., Hobson, V.J., Kubicek, L., Houghton, J.D., 2010. The biology and ecology of the ocean sunfish mola mola: a review of current knowledge and future research perspectives. *Reviews in Fish Biology and Fisheries* 20, 471–487. doi:[10.1007/s11160-009-9155-9](https://doi.org/10.1007/s11160-009-9155-9).
- <sup>24</sup> Provoost, P., Bosch, S., 2020. robis: R client to access data from the obis api. ocean biogeographic information system, intergovernmental oceanographic commission of unesco URL: <https://cran.r-project.org/package=robis>.
- R Core Team, 2018. R: A language and environment for statistical computing. vienna, austria: R foundation for statistical computing URL: <https://www.r-project.org/>.
- <sup>27</sup> Ramírez, F., Afán, I., Davis, L.S., Chiaradia, A., 2017. Climate impacts on global hot spots of marine biodiversity. *Science Advances* 3, e1601198. doi:[10.1126/sciadv.1601198](https://doi.org/10.1126/sciadv.1601198).
- <sup>30</sup> Sala, E., Mayorga, J., Bradley, D., Cabral, R.B., Atwood, T.B., Auber, A., Cheung, W., Costello, C., Ferretti, F., Friedlander, A.M., et al., 2021. Protecting the global ocean for biodiversity, food and climate. *Nature* 592, 397–402. doi:[10.1038/s41586-021-03371-z](https://doi.org/10.1038/s41586-021-03371-z).
- <sup>33</sup> Telenius, A., 2011. Biodiversity information goes public: Gbif at your service. *Nordic Journal of Botany* 29, 378–381. doi:[10.1111/j.1756-1051.2011.01167.x](https://doi.org/10.1111/j.1756-1051.2011.01167.x).

- 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, 1098–1101.  
3 doi:[10.1038/nature09329](https://doi.org/10.1038/nature09329).
- Troia, M.J., McManamay, R.A., 2016. Filling in the gaps: evaluating completeness and coverage of open-access biodiversity databases in the united states. *Ecology and evolution* 6, 4654–4669. doi:[10.1002/ece3.2225](https://doi.org/10.1002/ece3.2225).
- Troia, M.J., McManamay, R.A., 2017. Completeness and coverage of open-access freshwater fish distribution data in the united states. *Diversity and Distributions* 23, 1482–1498. doi:[10.1111/ddi.12637](https://doi.org/10.1111/ddi.12637).
- Tuomisto, H., 2011. Do we have a consistent terminology for species diversity? yes, if we choose to use it. *Oecologia* 167, 903–911. doi:[10.1007/s00442-011-2128-4](https://doi.org/10.1007/s00442-011-2128-4).
- 12 Turner, M.G., Calder, W.J., Cumming, G.S., Hughes, T.P., Jentsch, A., LaDeau, S.L., Lenton, T.M., Shuman, B.N., Turetsky, M.R., Ratajczak, Z., et al., 2020. Climate change, ecosystems and abrupt change: science priorities. *Philosophical Transactions of the Royal Society B* 375, 20190105. doi:[10.1098/rstb.2019.0105](https://doi.org/10.1098/rstb.2019.0105).
- 15 UNEP-WCMC, IUCN, 2022. Protected Planet: The World Database on Protected Areas (WDPA) [Online], January 2022, Cambridge, UK. Technical Report. URL: <https://www.protectedplanet.net>.
- 18 Visalli, M.E., Best, B.D., Cabral, R.B., Cheung, W.W., Clark, N.A., Garilao, C., Kaschner, K., Kesner-Reyes, K., Lam, V.W., Maxwell, S.M., et al., 2020. Data-driven approach for highlighting priority areas for protection in marine areas beyond national jurisdiction. *Marine Policy* 122, 103927. doi:[10.1016/j.marpol.2020.103927](https://doi.org/10.1016/j.marpol.2020.103927).
- 21 Webb, T.J., Vanden Berghe, E., O'Dor, R., 2010. Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PloS one* 5, e10223. doi:[10.1371/journal.pone.0010223](https://doi.org/10.1371/journal.pone.0010223).
- 24 Wickham, H., Francois, R., Henry, L., Müller, K., 2021. dplyr: A grammar of data manipulation. r package version 1.0.3. R Found. Stat. Comput., Vienna URL: <https://CRAN.R-project.org/package=dplyr>.
- 27 WORMS, 2022. World register of marine species database: Statistics. number of records in worms 11th april 2022 [online] URL: <http://www.marinespecies.org/>.
- 30 Yang, W., Ma, K., Kreft, H., 2013. Geographical sampling bias in a large distributional database and its effects on species richness–environment models. *Journal of Biogeography* 40, 1415–1426. doi:[10.1111/jbi.12108](https://doi.org/10.1111/jbi.12108).

- Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A., et al., 2007. How global is the global biodiversity information facility? *PloS one* 2, e1124. doi:[10.1371/journal.pone.0001124](https://doi.org/10.1371/journal.pone.0001124).
- Zhang, Y., Grassle, J.F., 2002. A portal for the ocean biogeographic information system. *Oceanologica Acta* 25, 193–197. doi:[10.1016/S0399-1784\(02\)01204-5](https://doi.org/10.1016/S0399-1784(02)01204-5).
- 6 Zizka, A., Carvalho, F.A., Calvente, A., Baez-Lizarazo, M.R., Cabral, A., Coelho, J.F.R., Colli-Silva, M., Fantinati, M.R., Fernandes, M.F., Ferreira-Araújo, T., et al., 2020. No one-size-fits-all solution to clean gbif. *PeerJ* 8, e9916. doi:[10.7717/peerj.9916](https://doi.org/10.7717/peerj.9916).

## Appendix A. The database

Table A.4 below shows the data loss for each criterion that we have used to clean our database. We downloaded 71,670,596 records from GBIG and OBIS. Only 820,004 records were useful for our analyses.

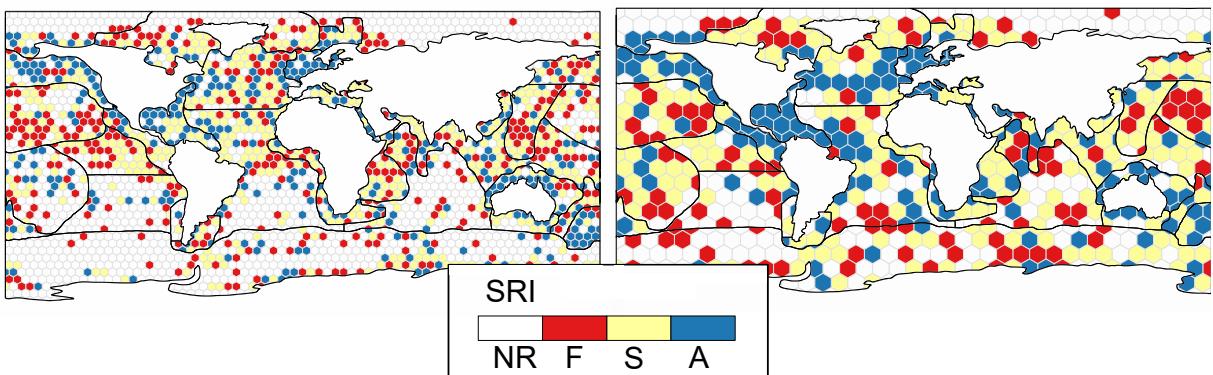
Database state	Number of records
Orignal records from GBIF and OBIS	71,670,596
Data curation (following Zizka et al. (2020))	5,380,439
Taxonomically filtered data	5,007,322
Deletion of data outside the native range	820,004

**Table A.4:** Criteria for filtering occurrence data from GBIF and OBIS using bioregions.

Files of the 10,371 marine fish species and their attributes (body size, habitat depth, and cultural value) from FishBase may be found in the GitHub project page of this manuscript: [http://github.com/vapizarro/stp\\_fishes](http://github.com/vapizarro/stp_fishes)

## Appendix B. Grids resolutions

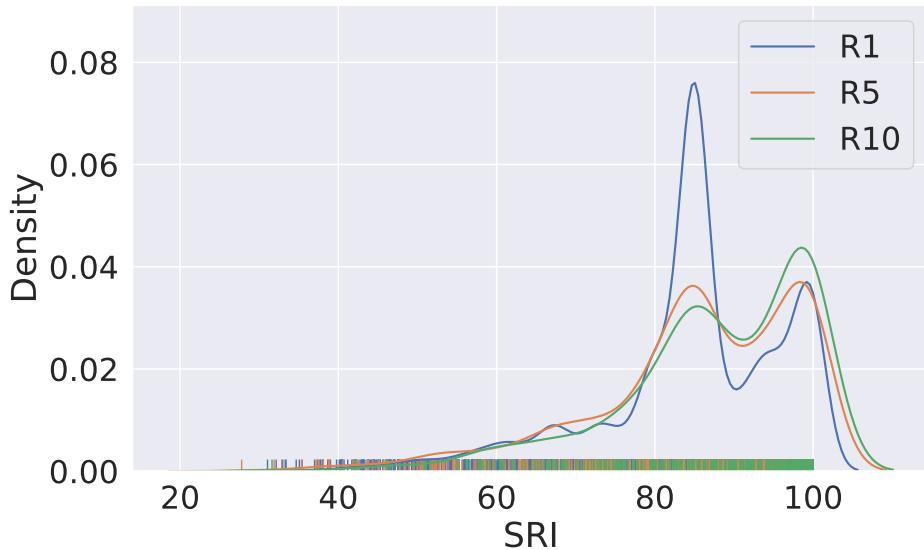
For spatial representation analysis we evaluated two additional spatial resolutions ( $5^\circ \times 5^\circ = 3,021$  cells, and  $10^\circ \times 10^\circ = 958$  cells). Table B.5 contains the results of this analysis for these grids. We have also mapped these results (see Figure B.7), to understand how the effect of spatial resolution on the evaluation of biodiversity macropatterns. Finally, we also plot the frequency of cells for each SRI category for the three grid sizes (R1= $1^\circ \times 1^\circ$ ; R5= $5^\circ \times 5^\circ$ ; R10= $10^\circ \times 10^\circ$ ) to understand how the data is distributed in our analyses (see Figure B.8)



**Fig. B.7:** Spatial representativeness index (SRI) mapping of cells of size: A= $5^\circ \times 5^\circ$ ; B= $10^\circ \times 10^\circ$ . The categorization of the cells corresponds to the level reached by the SRI, where SRI < 0.85: Amount of data Adequate for the representation of species richness (“A”); SRI=0.60-0.85: Amount of data can be considered Sufficient (“S”); SRI=0-0.60: Amount of records Few (“F”); and SRI = NA: cells with no records (“NR”).

## 15 Appendix C. Bioregions slopes

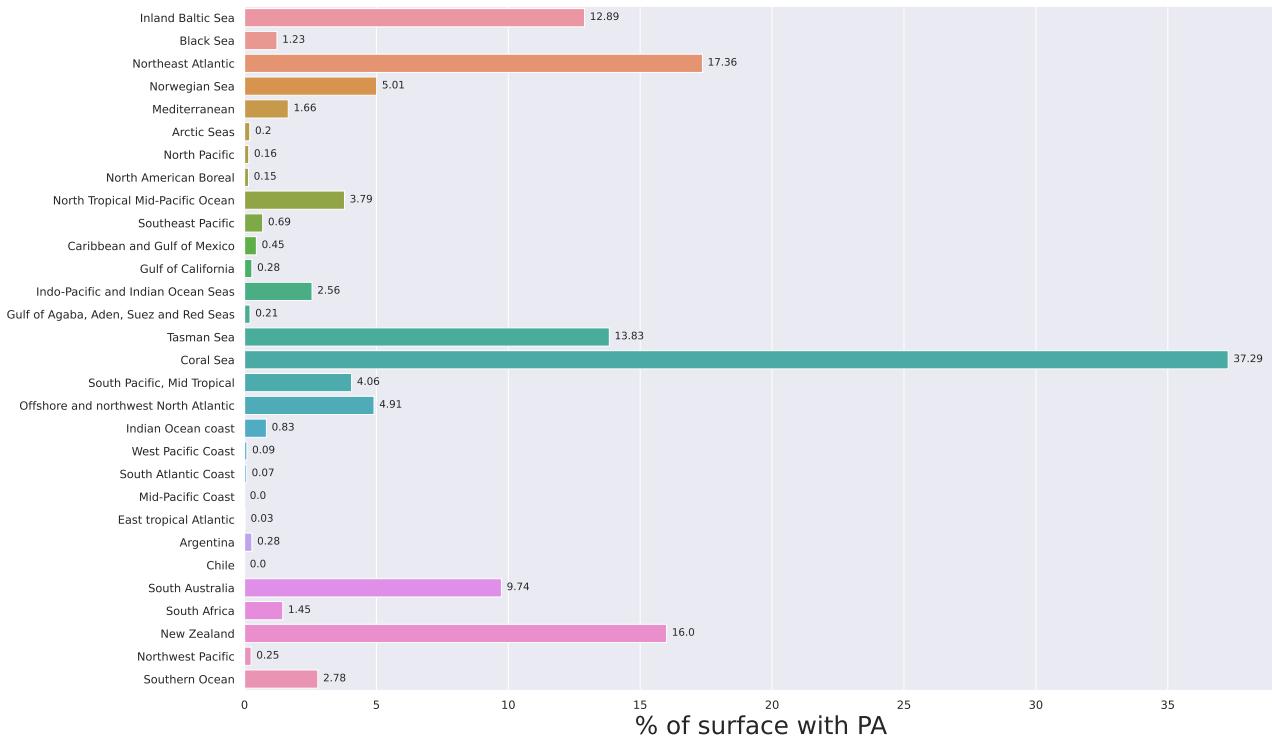
We evaluated the slopes of the last 10% of the accumulation curves of each bioregion in our temporal representation analysis. Table C.6 shows the result for each bioregion.



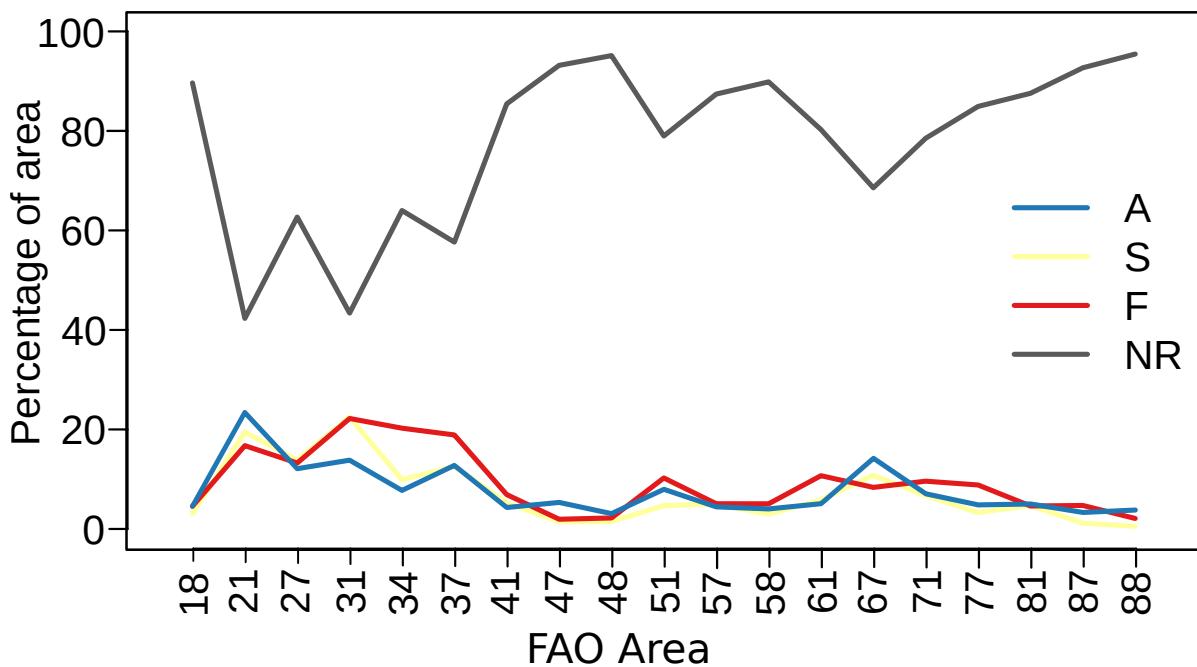
**Fig. B.8:** Density probability distribution of SRI in three grids of different sizes: R1 =  $1^\circ \times 1^\circ$  (blue line); R5 =  $5^\circ \times 5^\circ$  (red line); and R10 =  $10^\circ \times 10^\circ$  (yellow line).

#### Appendix D. Supplementary Material: GAP Analysis

We plotted the percentage of surface with marine protected areas of each bioregion (Fig 3 D.9), and the percentage of cells of each FAO Area for each category of SRI value (Fig D.10).



**Fig. D.9:** Percentage of surface with marine protected areas by bioregions.



**Fig. D.10:** Percentage of cells of each FAO Area for each category of SRI value. Amount of data *Adequate* for the representation of species richness (“A”); SRI=0.60-0.85; Amount of data can be considered *Sufficient* (“S”); SRI=0-0.60; Amount of records *Few* (“F”); and SRI = NA: cells with no records (“NR”).

ID	$5^\circ \times 5^\circ$				$10^\circ \times 10^\circ$			
	NR	F	S	A	NR	F	S	A
1	19.38	3.52	22.18	54.92	0.17	0.00	0.00	99.83
2	10.22	67.79	22.00	0.00	0.24	79.80	13.41	6.55
3	14.17	5.86	3.68	76.30	17.26	0.00	1.65	81.09
4	8.11	18.15	49.78	23.96	24.40	0.00	56.89	18.71
5	17.61	11.41	51.06	19.91	0.76	21.01	29.87	48.37
6	72.22	13.59	7.01	7.18	57.95	14.53	10.67	16.85
7	22.68	19.18	30.95	27.26	7.23	5.57	42.67	44.52
8	28.29	32.88	28.34	10.49	10.75	38.44	49.14	1.68
9	38.10	36.07	20.03	5.80	8.97	49.76	30.88	10.39
10	70.33	14.72	8.21	6.73	55.39	17.91	15.12	11.58
11	15.36	0.00	37.46	47.18	3.84	0.00	10.04	86.11
12	21.88	17.33	49.92	10.87	2.44	9.04	53.22	35.30
13	15.01	34.79	32.35	17.85	8.84	25.58	48.64	16.94
14	8.31	0.00	62.35	29.35	66.52	0.00	0.00	33.48
15	17.06	37.36	29.94	15.64	3.42	15.88	66.66	14.04
16	18.45	16.45	30.57	34.53	18.82	0.00	17.13	64.06
17	56.51	15.47	15.27	12.75	10.25	25.29	48.27	16.19
18	15.10	27.12	37.68	20.17	16.20	14.33	29.37	40.10
19	64.31	20.32	8.07	7.30	36.75	34.23	22.00	7.01
20	12.21	47.20	33.32	7.27	10.74	40.12	35.84	13.30
21	44.88	19.20	24.68	11.23	22.43	23.33	30.03	24.21
22	20.76	44.23	35.01	0.00	15.95	28.49	52.89	2.67
23	4.42	33.22	54.38	7.97	18.89	27.78	46.90	6.43
24	19.18	39.23	28.34	13.25	20.90	22.41	21.87	34.82
25	70.95	3.80	11.03	14.22	53.42	0.00	33.72	12.86
26	30.23	1.16	44.14	24.46	27.42	0.00	32.14	40.44
27	17.67	20.53	37.77	24.02	10.57	18.38	41.56	29.48
28	25.40	11.73	20.42	42.45	14.84	0.00	30.51	54.65
29	4.47	14.81	47.69	33.03	20.81	0.00	65.06	14.13
30	72.92	11.70	9.40	5.98	44.37	22.69	25.57	7.37

**Table B.5:** Surface area as a percentage of each bioregion (ID) for every SRI category for each of the three grid sizes ( $5^\circ \times 5^\circ$ ; R10= $10^\circ \times 10^\circ$ ). Amount of data *Adequate* for the representation of species richness (“A”); SRI=0.60-0.85: Amount of data can be considered *Sufficient* (“S”); SRI=0-0.60: Amount of records *Few* (“F”); and SRI = NA: cells with no records (“NR”).

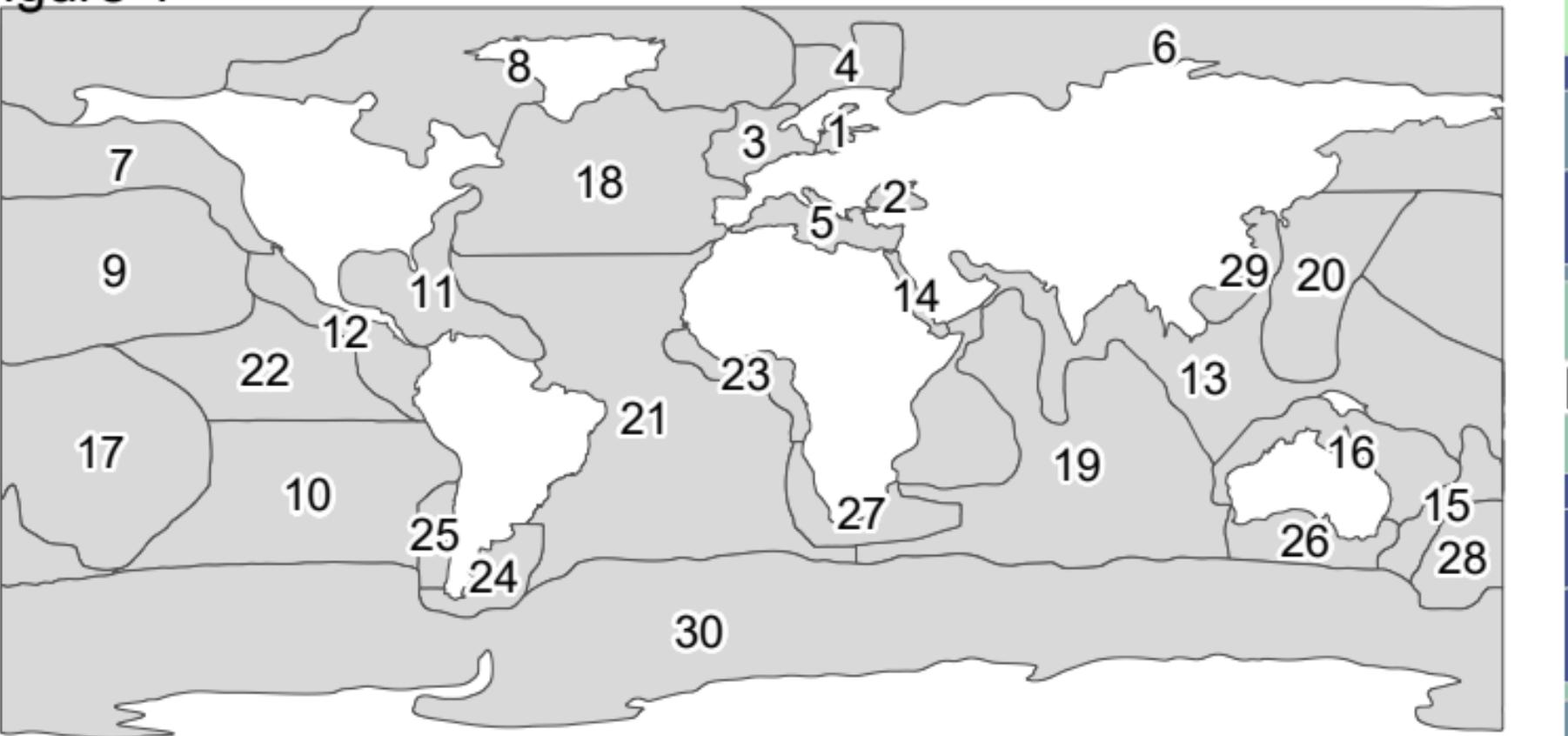
Bioregion	Slope
1	0.35
2	1.16
3	1.79
4	0.91
5	1.76
6	0.65
7	4.44
8	1.37
9	6.18
10	4.87
11	10.37
12	7.57
13	32.86
14	4.90
15	6.78
16	21.62
17	10.10
18	6.59
19	12.44
20	23.21
21	11.70
22	1.85
23	4.42
24	3.49
25	2.12
26	7.74
27	12.29
28	4.82
29	14.08
30	2.74

**Table C.6:** Final slope (10%) of the accumulation curves for each bioregion

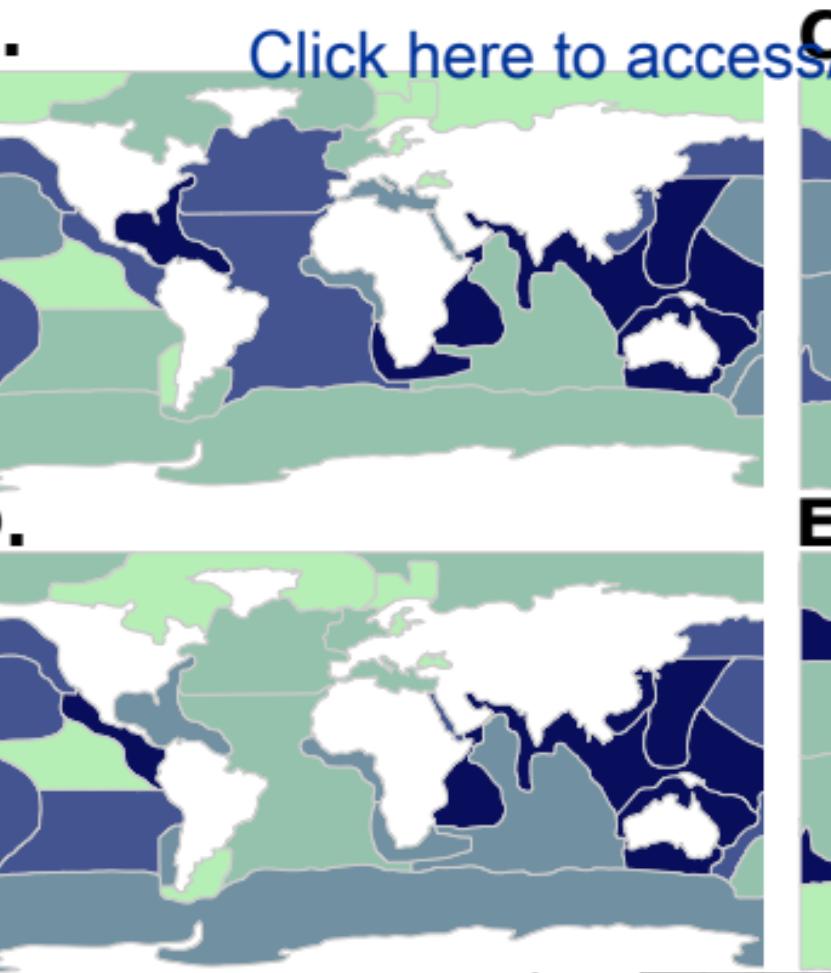


Click here to access/download  
**LaTeX Source File**  
Pizarro\_etAl\_Ecolnf.tex

**Figure 1**



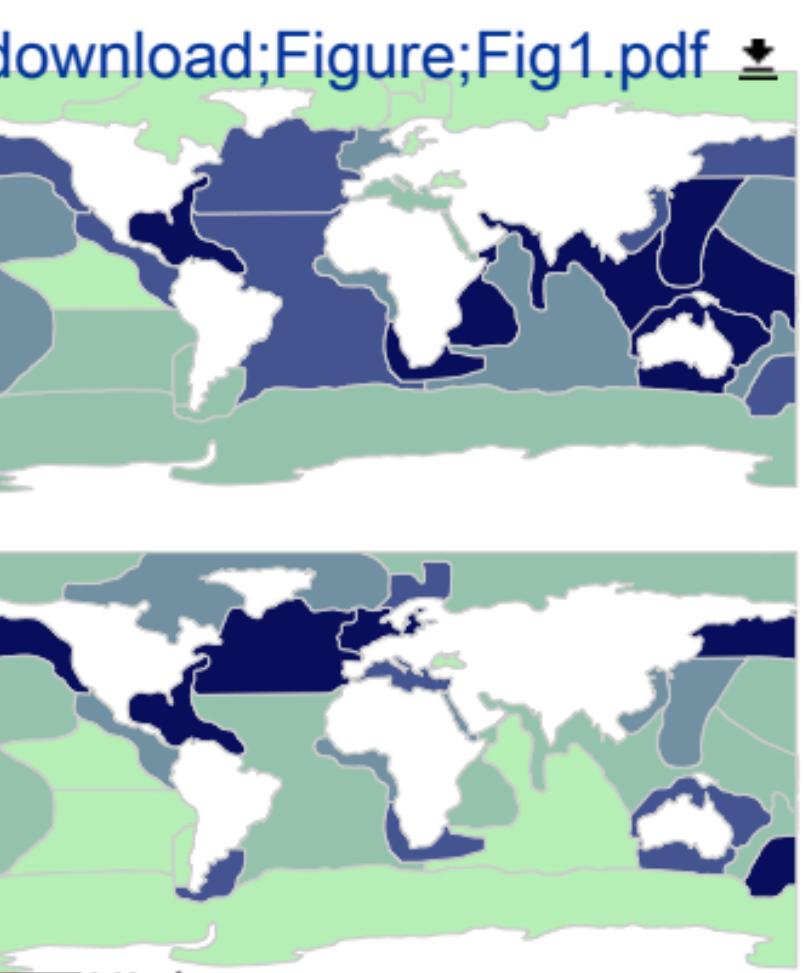
**B.**



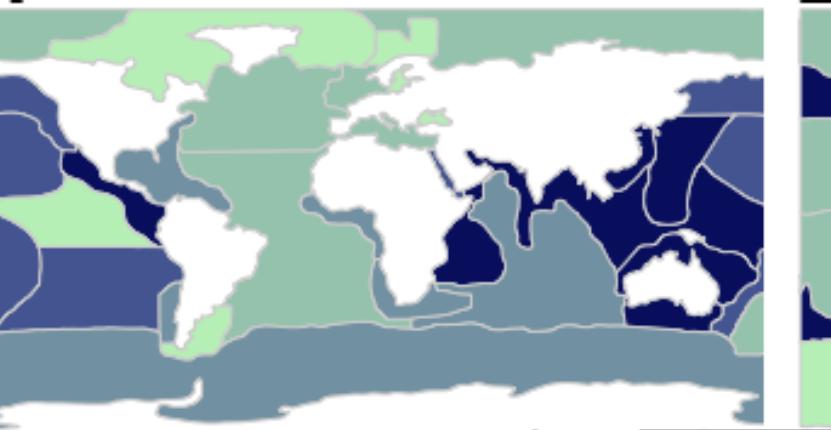
[Click here to access/download;Figure;Fig1.pdf](#)



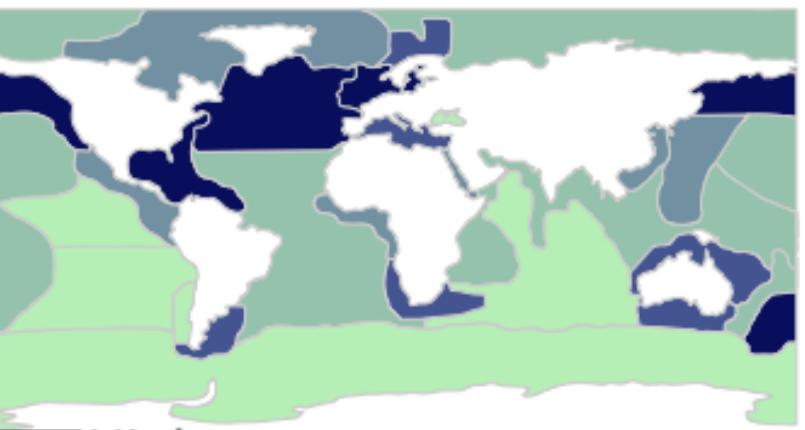
**C.**



**D.**

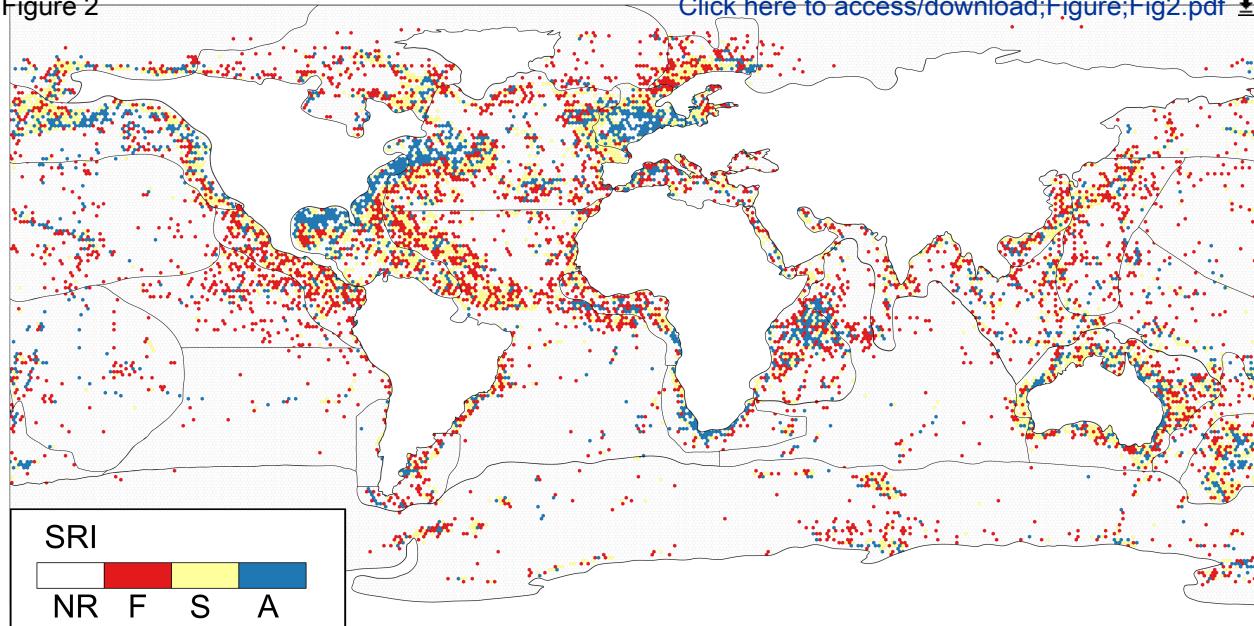


**E.**



Low High

Figure 2

[Click here to access/download;Figure;Fig2.pdf](#)

ID	NR	F	S	A	ID	NR	F	S	A
1	18.38	24.96	34.84	21.82	16	43.06	19.17	25.31	12.45
2	79.27	14.56	5.24	0.92	17	93.86	3.22	0.87	2.06
3	24.75	10.19	26.49	38.57	18	59.85	15.45	14.71	10.11
4	45.81	27.66	19.08	7.46	19	93.05	4.25	1.26	1.44
5	51.65	20.66	15.34	12.34	20	72.42	15.53	8.00	4.06
6	92.83	2.67	2.69	1.82	21	79.19	10.95	7.51	2.34
7	70.83	9.34	10.96	8.96	22	82.79	13.04	2.40	1.76
8	77.39	11.79	6.89	3.93	23	47.96	26.19	13.86	11.98
9	92.74	4.80	1.00	1.46	24	71.04	15.37	6.87	6.72
10	97.48	1.62	0.53	0.37	25	93.37	1.96	3.20	1.47
11	35.66	17.20	24.34	22.80	26	63.73	12.85	16.74	6.68
12	43.21	32.32	18.64	5.84	27	74.32	7.34	7.06	11.26
13	75.35	11.51	7.21	5.93	28	66.65	10.24	13.72	9.39
14	49.19	17.62	21.30	11.88	29	46.13	25.20	20.53	8.15
15	81.16	9.81	6.07	2.95	30	95.74	2.24	1.25	0.77

Figure 3

Click here to  
access/download;Figure;Fig

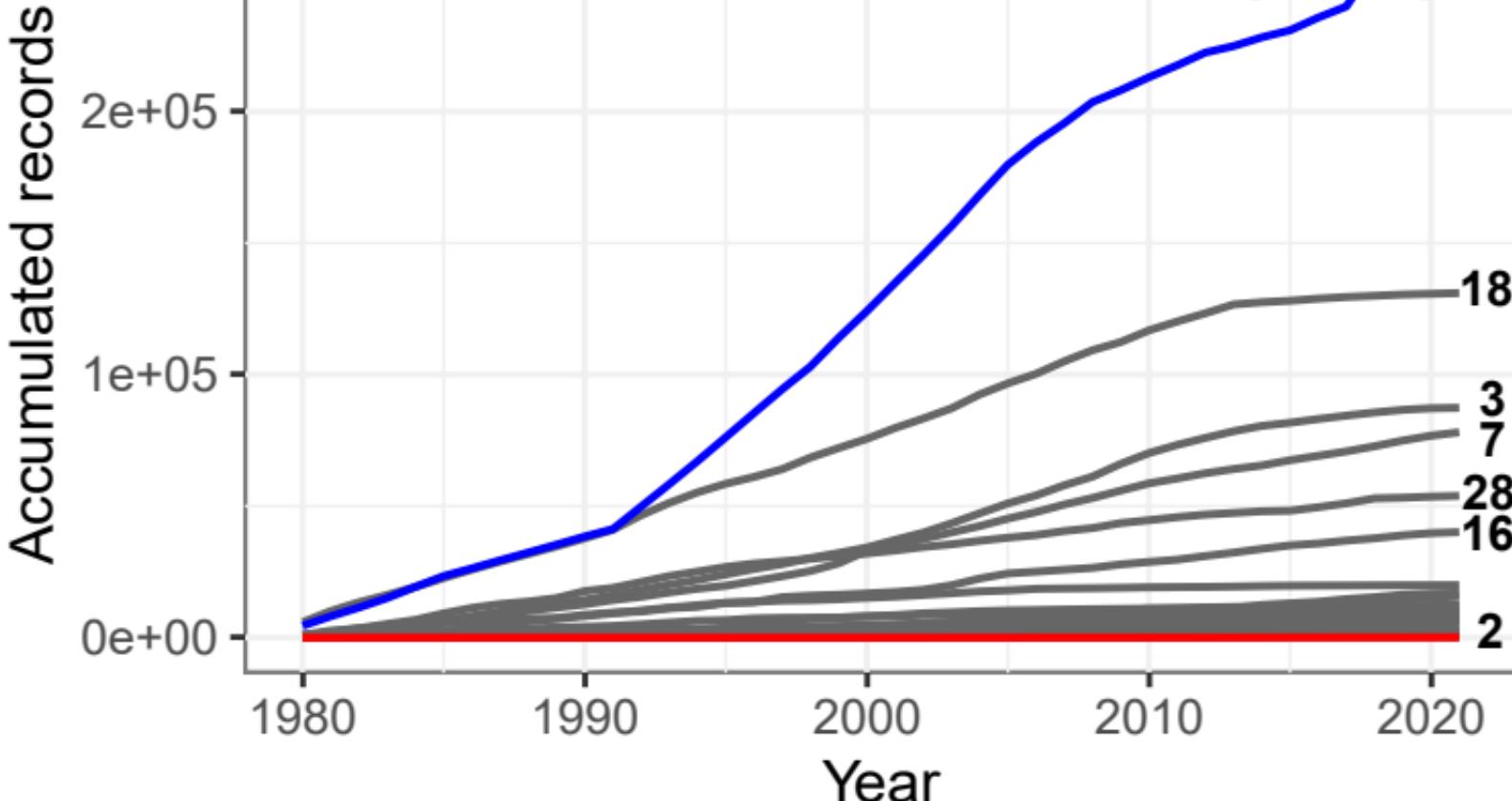
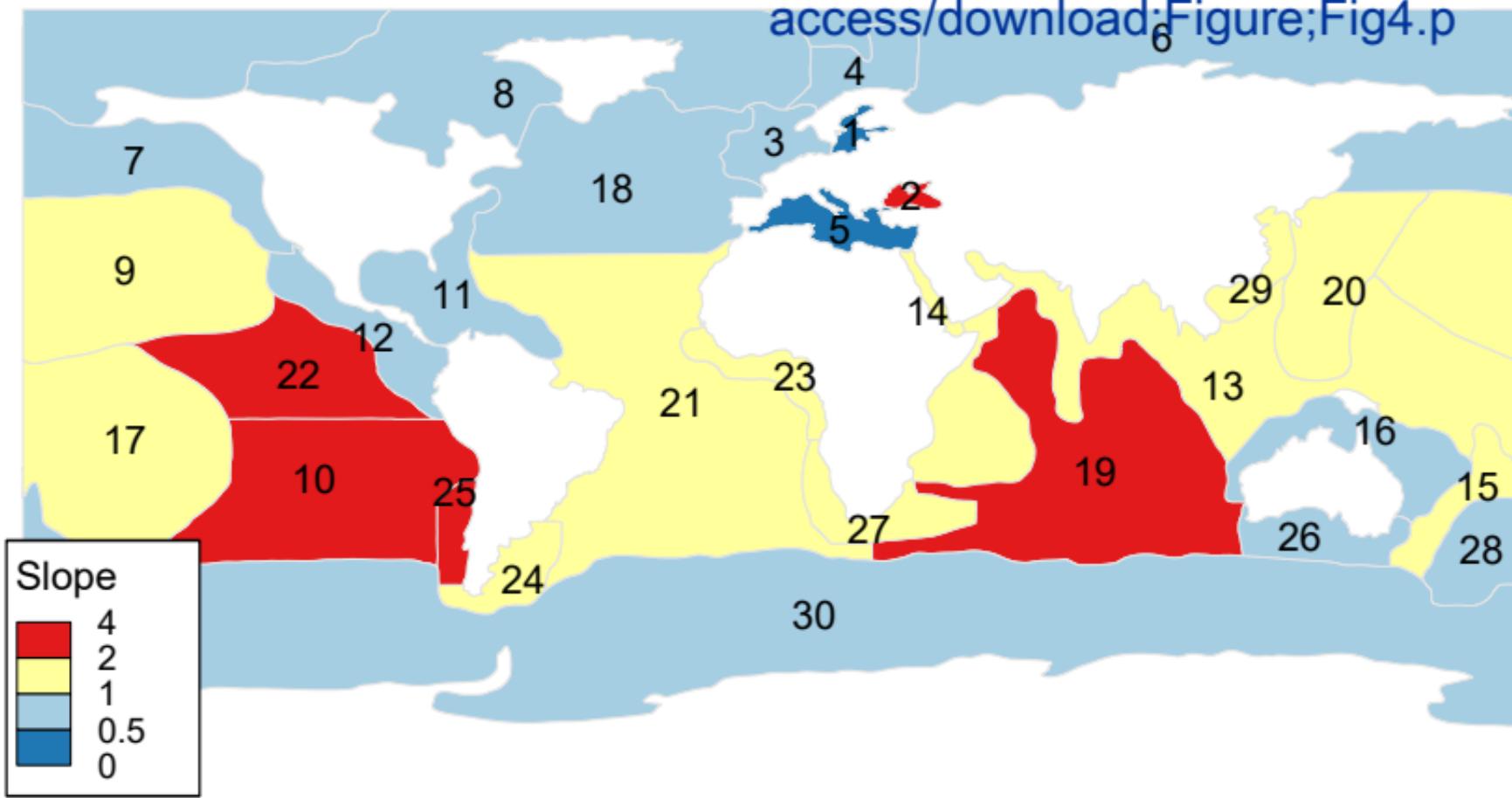
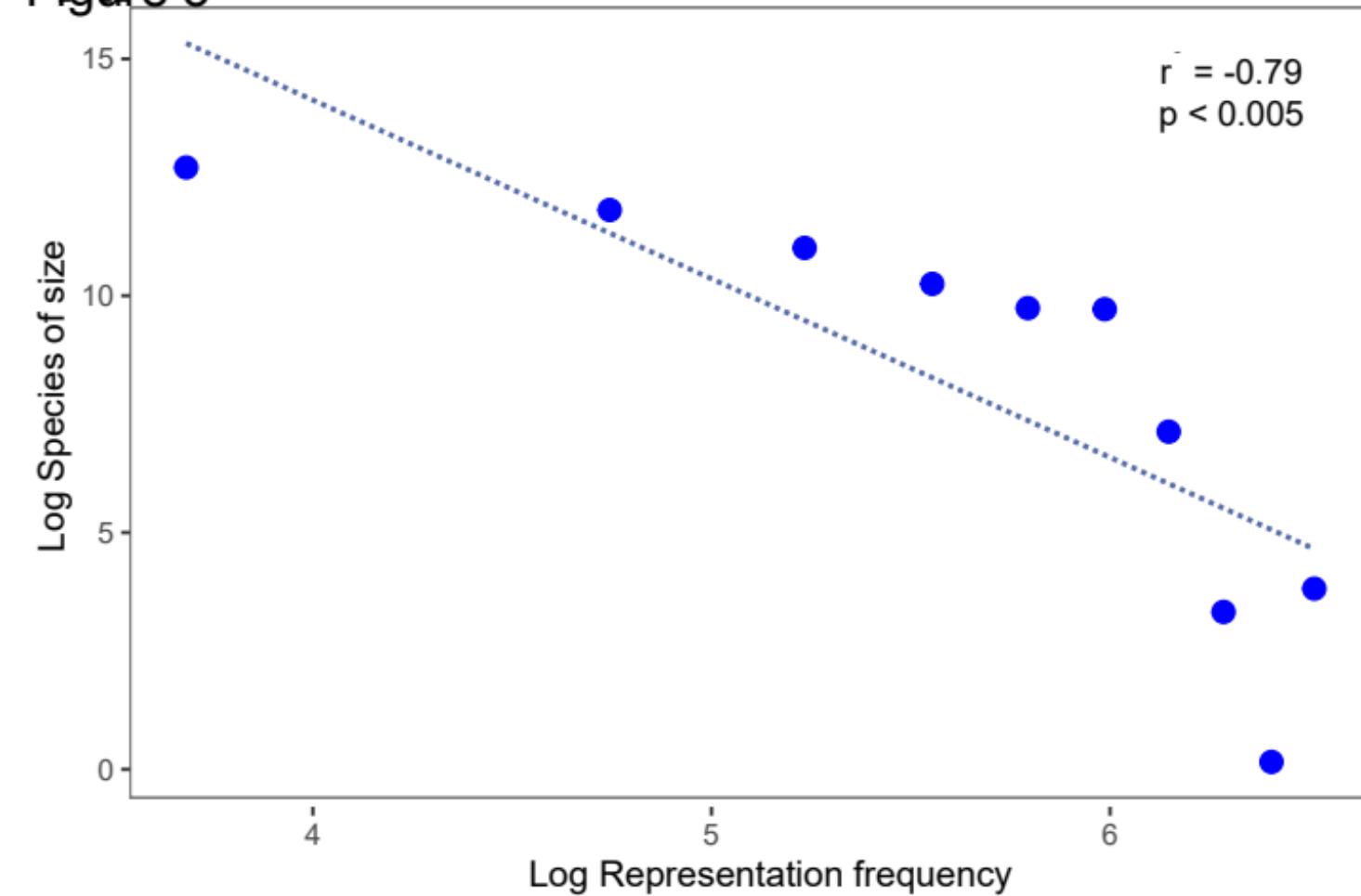


Figure 4

Click here to  
access/download:Figure;Fig4.p



**Figure 5**

[Click here to access/download;Figure;Fig5.pdf](#)

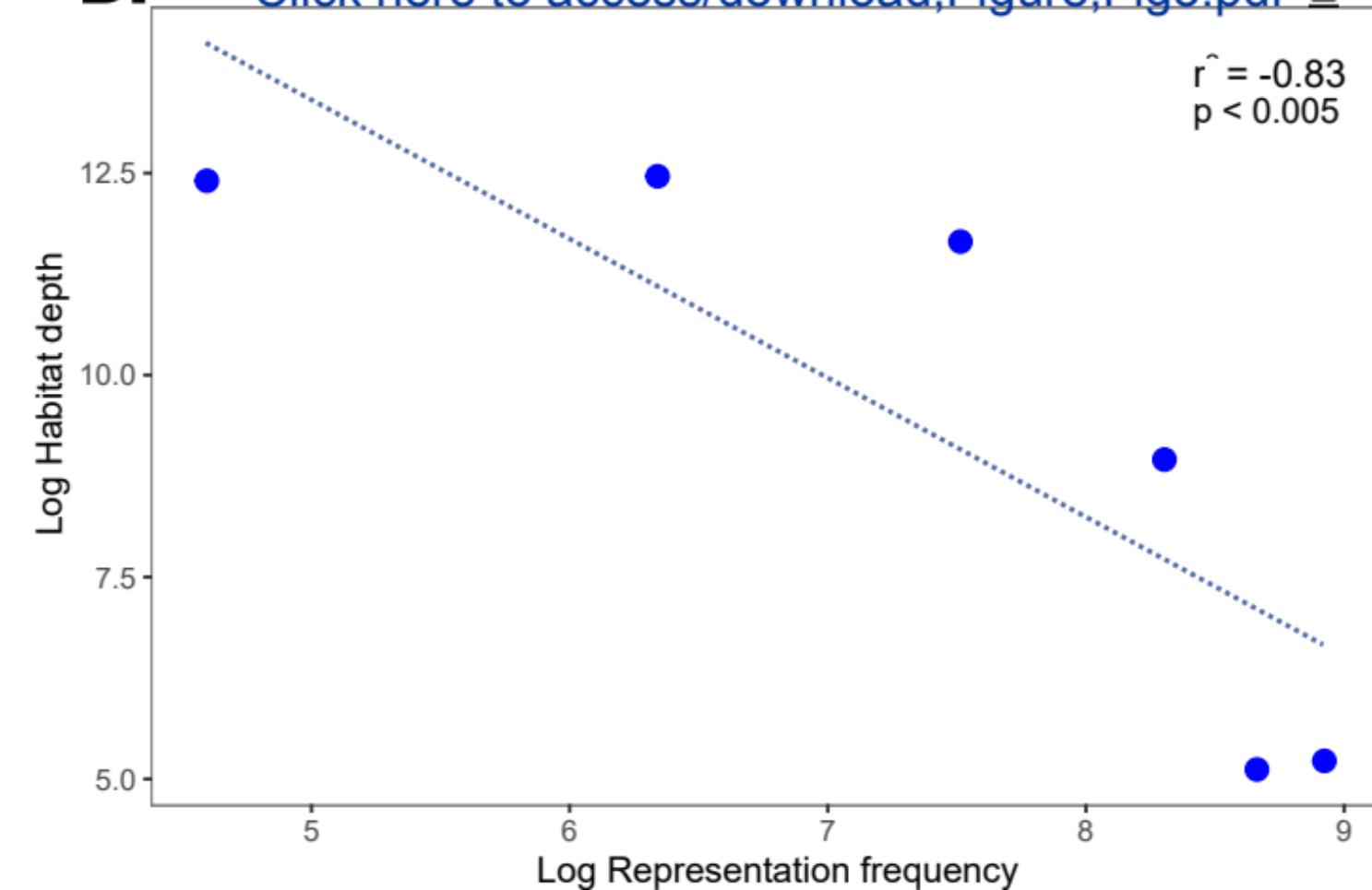
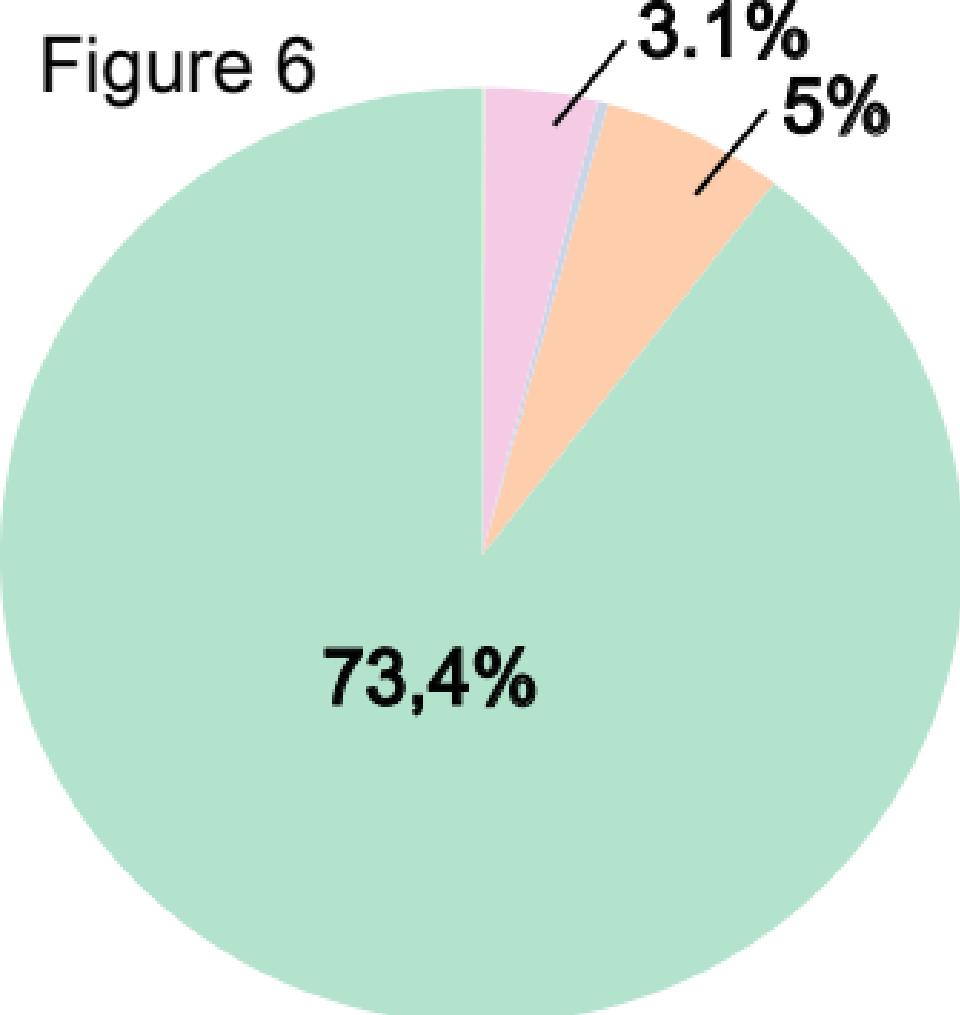


Figure 6



[Click here to access/download; Figure;](#) 

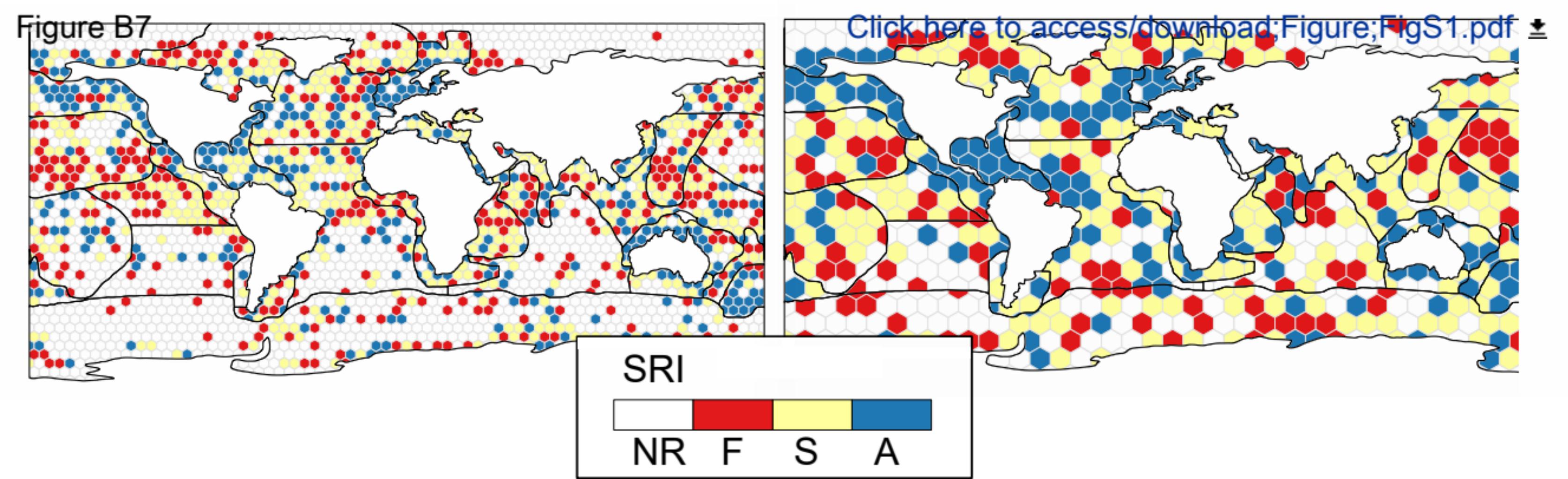


Figure B8

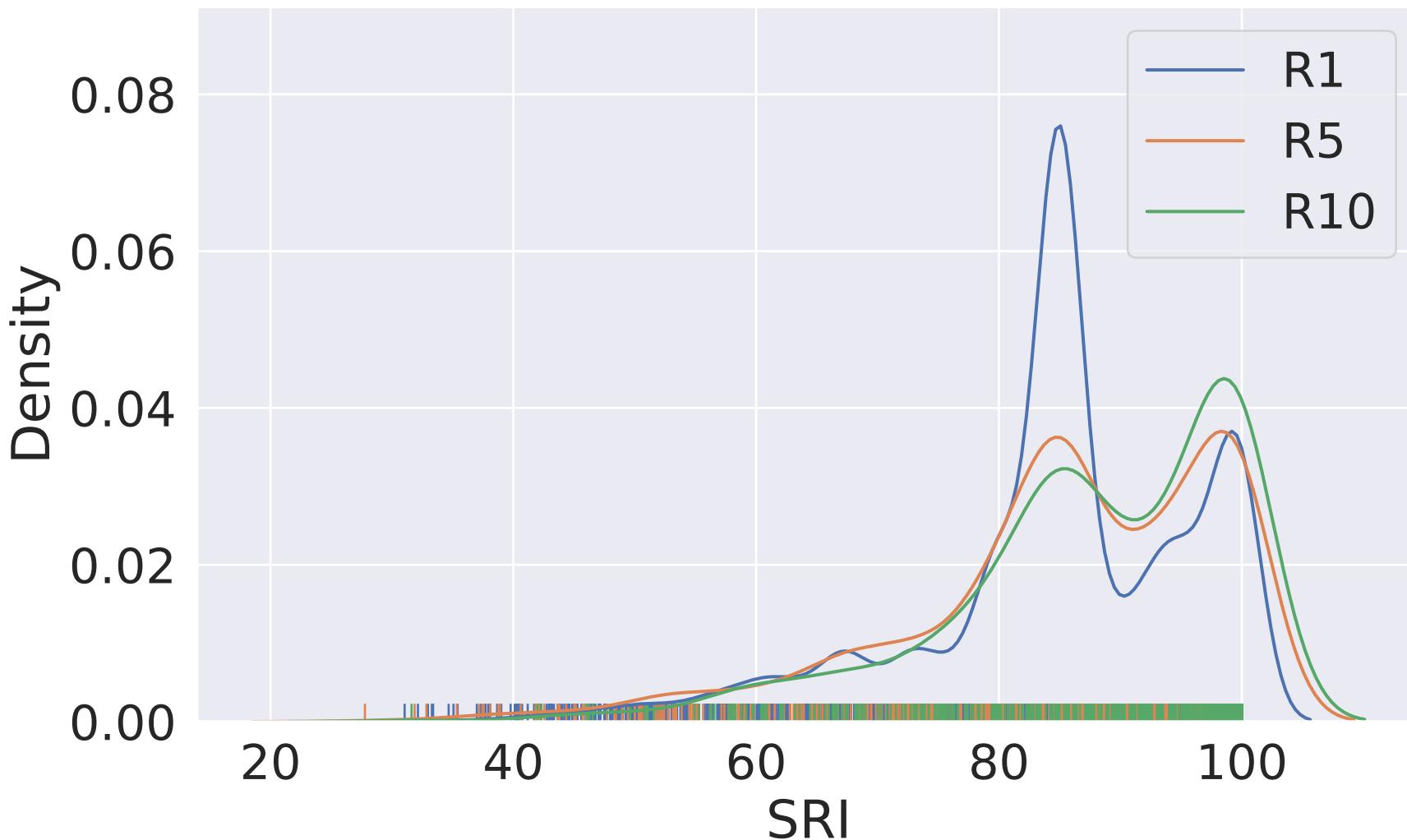
[Click here to access/download;Figure;FigS2.pdf](#)

Figure D9

[Click here to access/download;Figure;FigS3.pdf](#)