Aligning marine species range data to better serve science and conservation

Casey O'Hara[1], Jamie Afflerbach[1], Benjamin S. Halpern1, 2, 3

Author affiliation:

1. National Center for Ecological Analysis and Synthesis, 735 State Street Suite 300, Santa Barbara CA 93101
2. Bren School of Environmental Science and Management, University of California, Santa Barbara CA 93106
3. Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Rd, Ascot, West Berkshire SL5 7PY, United Kingdom

Corresponding author: Casey O'Hara, National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara CA 93101; (805) 892-2500; ohara@nceas.ucsb.edu

Classification: BIOLOGICAL SCIENCES: Ecology? Environmental Sciences? Sustainability Science?

Keywords: TBD

0.1 Abstract

Abstract. Provide an abstract of no more than 250 words on page 2 of the manuscript. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.

Species distribution data provide the foundation for a wide range of ecological research and conservation science and management, yet most species ranges remain unknown, and existing range maps have little overlap. In the ocean, two global species distribution datasets, produced by AquaMaps (REF) and the International Union for Conservation of Nature (IUCN) (REF), dominate our understanding of marine species ranges throughout the world's oceans. Together they represent 28,847 species, mostly from AquaMaps, with only 2,046 species overlapping. Here we examine differences in predicted species ranges between the two datasets, propose mechanistic causes and potential solutions for such differences, and explore the implications of these differences for management and conservation decisions. We find that IUCN maps often disregard bathymetry for depth-limited species, leading to predictions of species presence at unsuitable depths, and AquaMaps ranges for data-poor species often extrapolate presence far afield from known occurrences. We illustrate the implications of these differences by repeating two recent applications - the Ocean Health Index (Halpern et al. 2012, 2015) biodiversity goal and a global analysis of gaps in coverage of marine protected areas (Klein et al. 2015) - and find significantly different estimates of the status of global biodiversity and effectiveness of conservation depending on how the two datasets were prioritized. ***Creating a single, highly accurate dataset of global marine species ranges will be difficult. Understanding the implications of dataset differences for conservation planning and decision-making is essential.***

(223 words; 1593 char with spaces)

0.2 Significance

Significance Statement. Authors must submit a 120-word-maximum statement about the significance of their research paper written at a level understandable to an undergraduate-educated scientist outside their field of specialty. The primary goal of the Significance Statement is to explain the relevance of the work in broad context to a broad readership. The Significance Statement appears in the paper itself and is required for all research papers.

Section	Word #	Char # (w/spc)
Introduction	512 words	3411 char
Results/Discussion	802 words	5410 char
Implications (incomplete)	532 words	3334 char
Conclusions	390 words	2813 char
Methods (incomplete)	612 words	3889 char
total body	2848 words	18857 char

1 Introduction

Knowing where individuals of a species live is fundamental to the sciences of ecology, biogeography, and conservation, among many others. This knowledge provides foundational information for understanding species ranges and diversity, predicting species responses to human impacts and climate change, and managing and protecting species effectively. A rich literature tackles the many dimensions of these questions.

One major outcome of this body of science is the various compiled databases of species distribution maps. Two global-scale repositories predict marine species ranges throughout the world's oceans – AquaMaps (Kaschner et al. 2013) and International Union for Conservation of Nature (IUCN) (REF). These two spatial datasets have been used for a wide range of purposes, including assessing marine species status (Halpern et al. 2012, 2015 *both*; Selig et al. 2013 *both*), evaluating global biodiversity patterns (Coll et al. 2010 *AM*; Martin et al. 2014 *both and others*, Pimm et al. 2014 *IUCN, mostly focusing on terrestrial but some marine*, Kaschner et al. 2011), predicting range shifts (Molinos et al. 2015 *AquaMaps*, and setting conservation priorities (Klein et al. 2015 *AquaMaps*).

BSH: *I think Julie makes this comment below, but we may need to be able to substantiate this somehow. Is there a way to quickly figure out how many studies have used AM vs. IUCN?*

CCO: *What about age - AquaMaps is pretty recent (2006 at earliest); before that, would IUCN have been the dominant dataset for marine? Maybe we can track citations of AquaMaps, but citations of IUCN are generic - just cite the Red List - not referencing spatial information specifically, much less marine. Going by citation year field on the maps themselves, the oldest marine record I can find is 2004.*

Three foundational papers relating to AquaMaps (Kaschner et al. 2006, 2011; Ready et al. 2010) have been cited a total of 315 times.

The two datasets ostensibly describe the same information, but significant differences in methodology and intent could lead to dramatically different understandings of our marine ecosystems, with significant implications for policy and conservation recommendations. Importantly, biases in taxonomic or spatial coverage within a dataset could shift management and conservation actions away from places or species that are most in need. Inaccurate indications of presence or absence could lead to ineffective marine reserve systems and management plans (Rondinini et al. 2006) (Jetz 2008?).

To understand the implications of differences between the AquaMaps and IUCN datasets, we compare how each data source represents the global spatial and taxonomic distribution of species. Most notably, AquaMaps includes range maps for many more species (currently 22,889 species; 79.3% of total), such that most global analyses related to biodiversity to date have used AquaMaps (IUCN range map data exist for only 4,027 unique species). For the 2,013 species (7% of total) mapped in both datasets, we examine how well the maps align, determine several issues that lead to misalignment between predicted species distributions, and outline possible improvements.

We then reexamine two recent marine biodiversity studies - an assessment of the status of global biodiversity within the Ocean Health Index (Halpern et al. 2012, 2015) and a global analysis of gaps in protection afforded by marine protected areas (MPAs) (Klein et al. 2015) - as case studies to explore the implications of prioritizing one data set over the other. The results highlight the possible consequences of different data use decisions on our understanding of marine biodiversity status and protection.

---

2 Results and Discussion

2.1 How and why the datasets differ

The IUCN publishes species range maps, bundled by taxonomic groups, based on expert input on spatial boundaries of a given species' "limits of distribution" (IUCN 2015) - essentially a refined extent of occurrence, based on observation records and refined by expert understanding of the species' range and habitat preferences. In contrast, AquaMaps models species distribution based environmental preferences (e.g. temperature, depth, salinity) deduced from occurrence records, published species databases such as FishBase, and expert knowledge. The AquaMaps model overlays these environmental preferences atop a map of environmental attributes on a 0.5 degree grid, creating a global raster of probability of occurrence for each species (Kaschner et al. 2006, Ready et al. 2010)

The methodologies behind these datasets imply differences in prediction of species distribution due to errors of commission (falsely *(erroneously? less accusatory, but an annoying word)* indicating species presence) and omission (falsely indicating species absence). Geographic range data such as IUCN range maps generally include large commission errors, while predicted distribution models such as AquaMaps will likely include fewer commission errors but more omission errors. Each type of error bears different implications for conservation goals: commission errors can result in prioritizing areas not relevant to conservation goals, while omission errors may result in protected area networks that fail to include important habitat and range (Rondinini et al. 2006).

The two datasets have notably different taxonomic (Fig. 1A) and regional (Fig. 1B,C) coverage. IUCN-mapped species focus on tropical latitudes and away from the Atlantic and Eastern Pacific compared to AquaMaps-mapped species. These differences can be mitigated by using both datasets, but the underlying methodological differences complicate such direct comparisons.

To explore differences in species distribution and range between the two datasets, we plotted the distribution alignment (where on the map) against the ratio of extent (how much of the map) for each shared species (Fig. 2A). A general negative linear pattern emerges, suggesting that increasing similarity in extent correlates with decreasing distribution alignment. AquaMaps tends to extrapolate species ranges beyond known occurrences, such that the marginal range predicted by AquaMaps will fall in different locations than the marginal range predicted using IUCN methodology. For species with dissimilar extents, predicted distribution for the smaller range can more easily fall within the generous bounds of the larger range. For species with increasingly similar extents, differences in methodology become more difficult to "hide," and the distribution alignment generally becomes poorer.

By dividing the map-paired species into quadrants, we highlight different categories of relationships that in turn help further explain the general pattern. The upper right quadrant includes species (n = 399) whose described ranges agree in both spatial distribution and extent. These species tend to be well-studied and include wide-ranging pelagic organisms such as marine mammals, tunas, and billfishes (Fig. 2B). This result is not surprising, as species with very large ranges are likely to be more aligned regardless of methodology simply because they exist nearly everywhere.

The extent-mismatched ranges contained in the upper left quadrant (n = 624) include many species whose spatial distribution is similar, but where the IUCN range is notably larger, often extending into deeper water. For example, corals dominate this quadrant (n = 221; 35.4% of all species in this quadrant), and IUCN range maps tend to extend corals into waters beyond their preferred depths, likely introducing errors of commission. Ocean depth is explicitly included in AquaMaps models, while depth is frequently overlooked as a factor in IUCN range maps. Simply clipping IUCN range maps to known depth preferences would resolve many of these mismatches.

Species found in the lower right quadrant (n = 624) often represent cases of "two wrongs make a right." For these species, IUCN ranges frequently overextend into unsuitable depths, as in the case of many upper left quadrant species, while at the same time AquaMaps ranges often aggressively extrapolate presence into locations where IUCN predicts absence, introducing additional commission errors. Consequently, area ratios are close to 100%, though similar extents are meaningless when the distributions are poorly aligned.

The most vexing cases are in the lower left quadrant (n = 399), where neither distribution nor extent match well. Data-poor species are more common in this quadrant; indeed, the median number of species occurrence records (summing (averaging?) OBIS and GBIF occurrences) for this quadrant is NA, compared to a median of for species across the other three quadrants. When extrapolating from limited observations, the AquaMaps model often predicts species presence well beyond known occurrences, introducing commission errors; at the same time, IUCN range maps generally target known occurrences, possibly introducing omission errors for data-limited species.

---

3 Implications

Clear and significant differences in species range descriptions between AquaMaps and IUCN range maps, due to method-driven differences in errors of commission and omission. To examine the implications of these differences, we replicated two recent studies, varying only the prioritization of one data set over the other.

Case Study: The Ocean Health Index

The global Ocean Health Index (OHI) (Halpern et al. 2012, 2015), a composite index comprising ten sustainable benefits provided by a healthy ocean, uses species spatial distribution data and IUCN Red List conservation status to calculate biodiversity status (scored from zero to 100) for each of the world's 221 exclusive economic zones. To maximize the number of represented species, OHI gleans spatial distributions from both IUCN and AquaMaps datasets, prioritizing IUCN data for the 2013 species included in both sources. OHI currently uses a probability threshold of 40% to determine species presence for AquaMaps data.

We recalculated the OHI species status score under several scenarios to observe the impact of toggling the prioritized data set from IUCN to AquaMaps, and toggling the AquaMaps presence threshold from 40% to 0% (Fig. 3). Given that only 2013 species overlap out of 28847, it is surprising that changing the priority for overlapping species from IUCN maps to AquaMaps would result in such large score shifts as seen in scenarios 2 and 3. While many country scores showed only small variation across all three scenarios, select countries gained as many as 5 points while others dropped up to 12 points. Significantly, the global score decreased by two full points in scenario 3, indicating that an arbitrary change in how the two datasets are combined results in a more dire assessment of species conservation status.

Case Study: MPA Gap Analysis

Klein et al. (2015) compared the global distribution of species to the global distribution of marine protected areas to assess how well the MPAs protect key species and to identify which species fall through gaps in protection. The study relied on the AquaMaps database, using a probability threshold of 50% or greater, to determine species presence, and the World Database of Protected Areas to define zones of marine protection. The study determined that currently established MPAs (IUCN categories I through IV overlapping with marine areas) are woefully inadequate in protecting marine biodiversity; 90.5% of species have less than 5% of their overall range represented within MPAs and 1.4% of species have no protection at all ("gap" species).

(Figure 4) displays the results of applying this methodology under several different scenarios. When comparing the conclusions based upon an IUCN-only scenario to an AquaMaps-only scenario (using 2015 data and a 0% threshold sets the most meaningful comparison) we found a five-fold increase in the proportion of gap species (6.4% vs 1.2%) and dramatically larger proportion of species with less than 2% of their range protected (73.2% vs. 47.7%). However, this comparison also indicates a larger proportion of well-protected species with greater than 10% of range protected (2.9% vs. 1.5%). Conclusions drawn from each of these datasets would paint dramatically different pictures of the protection afforded by our current global MPA network.

---

4 Conclusions

AquaMaps and IUCN range maps provide important spatial information on marine species ranges, and show reasonable agreement for many well-studied species, but illustrate uncertainty in our understanding of spatial distribution for many others. Identifying and addressing differences in these datasets will increase their utility for ecology research and conservation actions. We have identified several likely drivers of commission errors and omission errors between these datasets, pointing toward improvements to better align the predictions of species distributions.

Suggestions for data providers:

- The IUCN Red List instructional materials for mapping marine species suggest that "bathymetry can be used to delineate species' ranges limited by depth in the same way as elevation is used for terrestrial species" (IUCN 2015). Unfortunately, this recommendation appears to be applied inconsistently. Especially for demersal communities dependent on photosynthesis, bathymetry should be a primary consideration; even a simple cutoff at the photosynthetic limit of 200 meters would minimize a likely source of commission errors.
- The AquaMaps model is a powerful tool for quickly estimating species distribution based on limited information; however, the model's output is only as good as the input data, which for marine species is often sorely limited. AquaMaps encourages experts to review and comment on the predicted species distribution maps; lessons from the small sample of reviewed maps could be generalized, for example to other species in the same taxonomic group, to further refine the predictions. This is especially important for data-poor species.
- In addition to environmental preferences and conditions, AquaMaps mapping parameters include area restrictions to help limit over-extrapolation of AquaMaps models. Currently area restrictions are denoted by UN Food and Agriculture Organisation (FAO) fisheries reporting areas (Ready et al. 2010); however, area restrictions based on biogeographical criteria, e.g. Marine Ecoregions of the World (Spalding et al. 2007), would likely provide better resolution and predictive power, especially for data-poor species.

Suggestions for data users:

- For depth-limited species and taxa, clipping IUCN range polygons to a reasonable bathymetry contour can reduce commission errors due to overprediction of species presence into unsuitable habitats.
- Using both datasets together can increase the taxonomic and spatial breadth of coverage, as long as the differences between the datasets can be reasonably minimized. Consider:

- An AquaMaps presence threshold of 0% most closely approximates the "limits of distribution" defined by IUCN range maps.
- Additional filters on AquaMaps distribution criteria, such as expert review status or the number of occurrence cells used to generate the species model, can help avoid overextrapolation for data poor species.

(do we need a wrap-it-up paragraph here? I put that more at the top of the conclusions section... could shift it down here, but probably not needed in both places)

---

5 Methods

Comparison of taxonomic and regional distribution: To examine the overall taxonomic distribution across the spatial datasets (Fig. 1A), we grouped species by taxonomic class and data source (IUCN, AquaMaps, or both), and examined the proportion of each class represented in each data source category.

To compare the spatial representation of the two datasets directly, we rasterized the IUCN species polygons to the same 0.5° grid as the AquaMaps species maps; species presence within a grid cell was determined by any non-zero overlap of a species polygon with the cell, and species richness per cell was simply the count of the species present. For the AquaMaps dataset, we determined per-cell species richness by counting all species with non-zero probability of occurrence, to best approximate the "extent of occurrence" generally indicated by IUCN maps. We represented relative distribution of species richness for each dataset by plotting average species count against latitude and longitude (Fig. 1B, 1C).

Comparison of paired maps: Using genus and species binomials as a matching key, we selected the subset of marine species that have range maps in both IUCN and AquaMaps current native distribution. To avoid double-counting, we removed subpopulations and species aliases. We determined species presence within each spatial cell using the same criteria as outlined above.

Overlaying paired distribution maps for a given species, we defined and calculated *distribution alignment* and *extent alignment* and plotted these in Fig. 2:

$$\alpha_{dist} = \frac{A_{small \cap large}}{A_{large}} * 100\%$$

$$\alpha_{area} = \frac{A_{small}}{A_{large}} * 100\%$$

We visually inspected a random selection of paired distribution maps from each quadrant to identify possible mechanistic causes of misalignment. To verify that IUCN predicted unsuitable habitat for depth limited species, we used QGIS (REF) to overlay a selection of IUCN and AquaMaps maps with a 200 meter bathymetry contour.

Methods for OHI case study: Using methods and supplemental materials from OHI (Halpern et al. 2012, 2015), we modified the original code for OHI 2015 Species status (SPP) (REF), allowing for flexibility in prioritized data source and AquaMaps presence threshold. We ran the SPP code three times, prioritizing IUCN over AquaMaps for a 0% threshold, and prioritizing AquaMaps over IUCN for both a 40% and 0% threshold. We compared each of these to the output of the published OHI 2015 SPP model (prioritizing IUCN over AquaMaps at a 40% threshold).

Methods for MPA Gap Analysis case study: Based upon the methods described in Klein et al. 2015, we reconstructed the analysis using the subset of protected areas (WDPA 2014) spatially covering a marine area and classified as IUCN I-IV. The WDPA polygons and marine polygons were rasterized to 0.01° and then aggregated to AquaMaps native 0.5° cells, to calculate proportion of protected area and marine within each cell. After verifying our results using the 2014 AquaMaps dataset, we updated the analysis using the 2015 AquaMaps dataset at a presence threshold of 50% (to compare to Klein et al. directly) and 0% (to better compare with IUCN spatial data). To analyze MPA coverage against IUCN spatial data, we extracted IUCN polygon weights per 0.5° cell for each species and compared against the protected area raster. Finally, we combined AquaMaps data (at 0% threshold) and IUCN data, using AquaMaps for the 2013 overlapping species and again compared against the protected area raster.

All processing was completed in R (REF), and all code and intermediate data are available on GitHub (REF?) at https://github.com/OHI-Science/IUCN-AquaMaps.

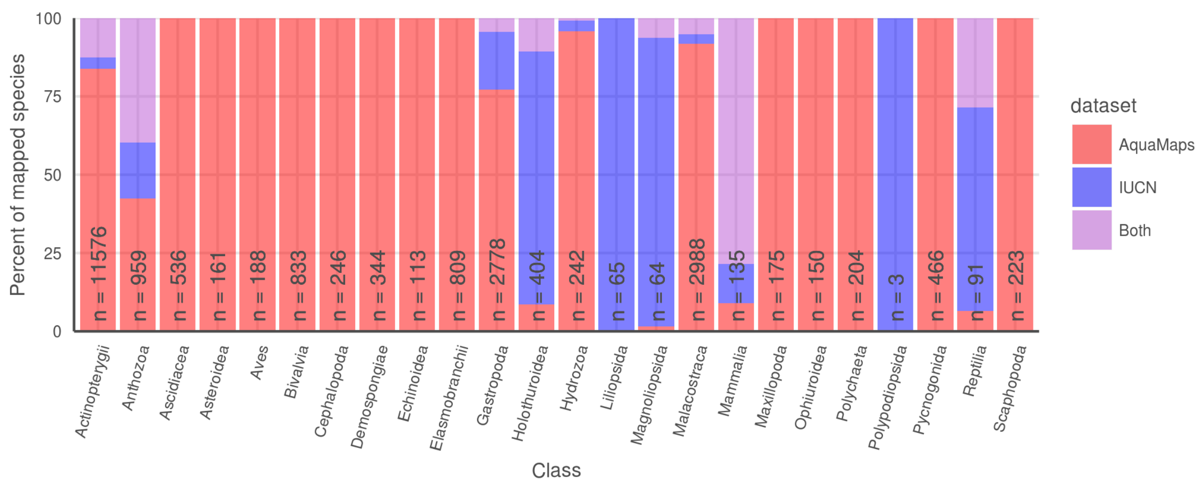
---

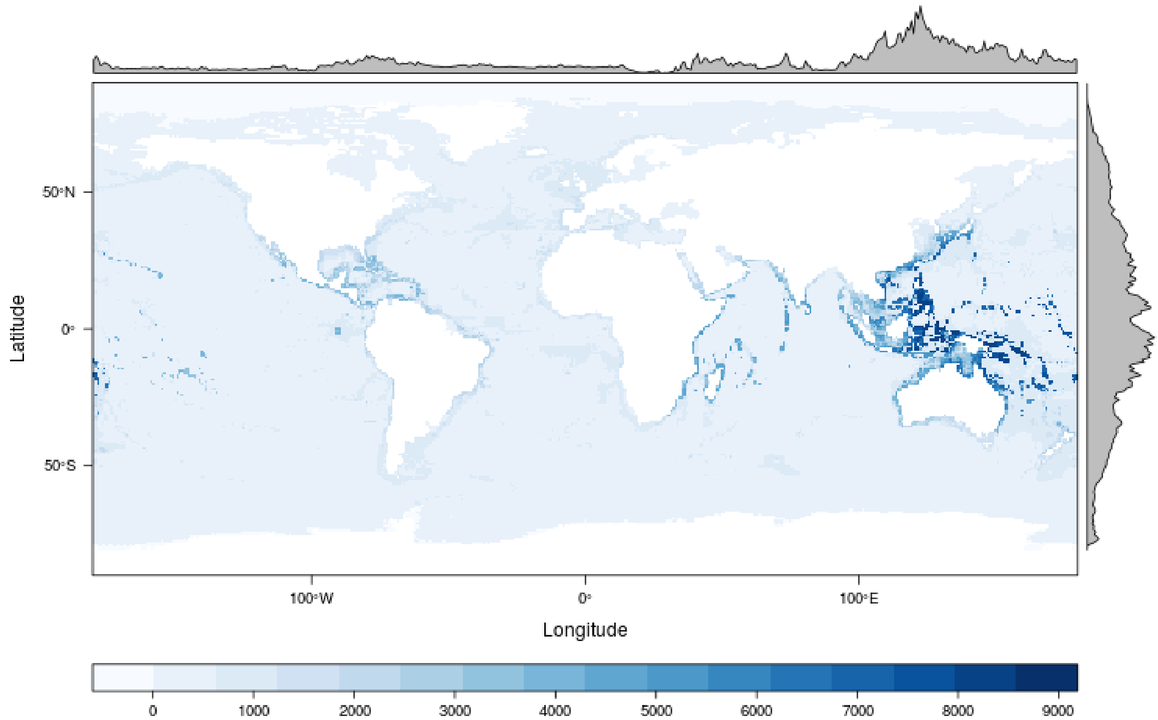
6 Figures and captions

still need work on captions; then separate captions from images

---

6.0.1 Fig. 1




50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

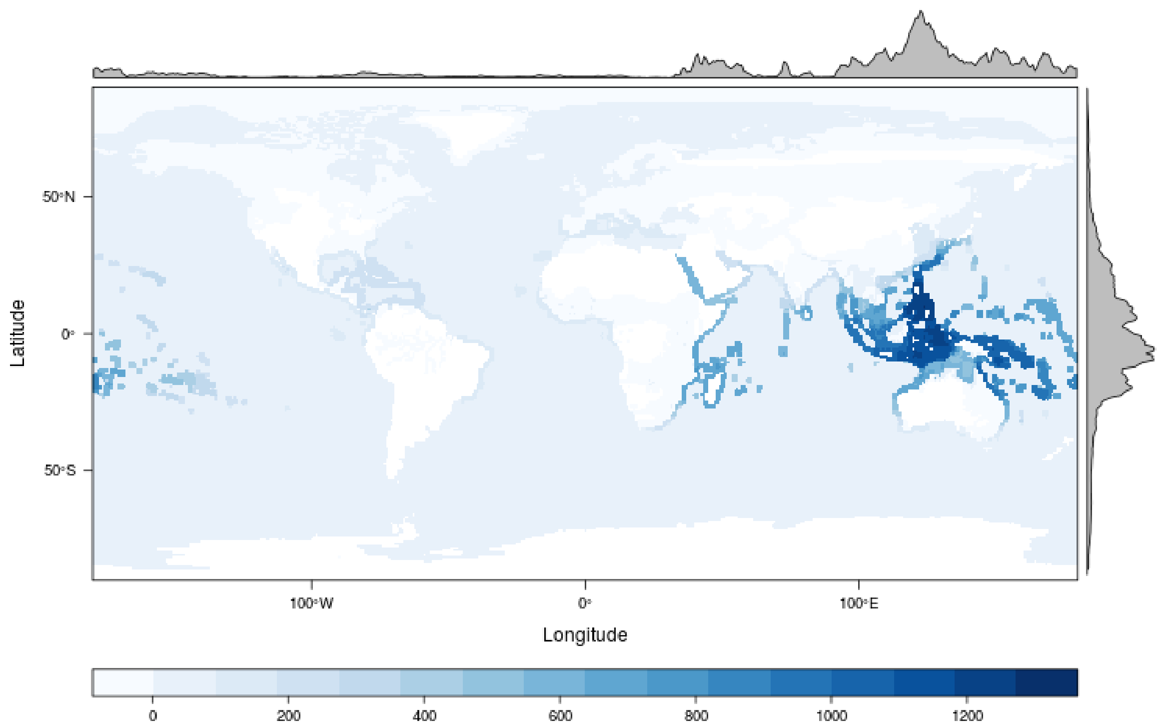


Fig. 1. (A) Number and proportion of species by taxa included in each dataset. AquaMaps encompasses a broader range of taxa than IUCN, while IUCN focuses on comprehensively assessing select taxonomic groups, typically at the level of order or family. Overlapping species are dominated by bony fishes (896 species, primarily tropical taxa) and corals (362 species). (B, C) Global marine species richness according to (B) AquaMaps dataset and (C) IUCN dataset. The margin frequency plots show relative species count per cell at each latitude and longitude; while both datasets peak in tropical latitudes near the equator, the frequency for IUCN maps drops quickly beyond 30°N and 30°S, while the frequency for AquaMaps remains robust well into temperate latitudes. The longitude frequency plots show a slight bias in the IUCN dataset away from the Atlantic and eastern Pacific compared to AquaMaps.

Fig. 2. (A) Distribution alignment (overlap of smaller range within larger) versus extent ratio (the ratio of smaller range area to the larger range area) for 2013 species included in both IUCN and AquaMaps datasets. The upper right quadrant (quadrant 1) comprises species whose maps largely agree (better than median value) in both spatial distribution and the extent of described ranges (n = 398; 19.5 %). The upper left quadrant (quadrant 2) comprises species whose maps agree well in distribution, but disagree in extent (n = 625; 30.5 %). The lower right quadrant (quadrant 3) includes species for which the paired maps generally agree in range extent, but disagree on where those ranges occur (n = 624; 30.5 %). The lower left quadrant (quadrant 4) indicates species for which the map pairs agree poorly in both area and distribution (n = 399; 19.5 %).
(B) Alignment breakdown of paired-map species by taxonomic group.

6.0.2 Figure 3

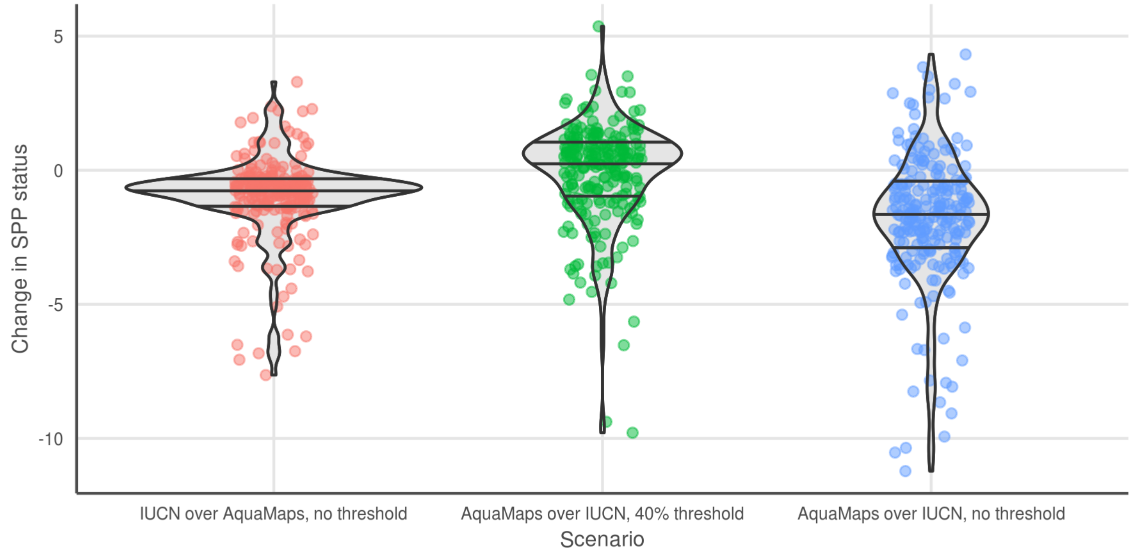


Fig. 3. Change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios. Scenario 1 shows the effect of reducing the probability threshold for AquaMaps presence to more accurately track the definition of IUCN "limits of distribution." Reducing the threshold increases the apparent range of a species; the slight decrease in average score suggests increased spatial representation of threatened species. Scenario 2 shows the effect of prioritizing AquaMaps data over IUCN, while maintaining the 40% presence threshold. In general, AquaMaps ranges are smaller than IUCN ranges, so most overlapping species will see a decrease in represented range. The slight bump in mean score may indicate a small increase in spatial representation of low-risk species, a small decrease in spatial representation of high-risk species, or a combination of both. Scenario 3 shows the effect of prioritizing AquaMaps data over IUCN, and simultaneously eliminating the presence threshold. The zero threshold in scenario 3 drives a decrease in scores relative to scenario 2. The large decrease indicates that a zero threshold greatly increases the spatial representation of high-risk species.

NOTE: This figure still needs to be recreated using the latest and greatest data! *currently they use data generated in November or so, in ohiprep.*

6.0.3 Figure 4

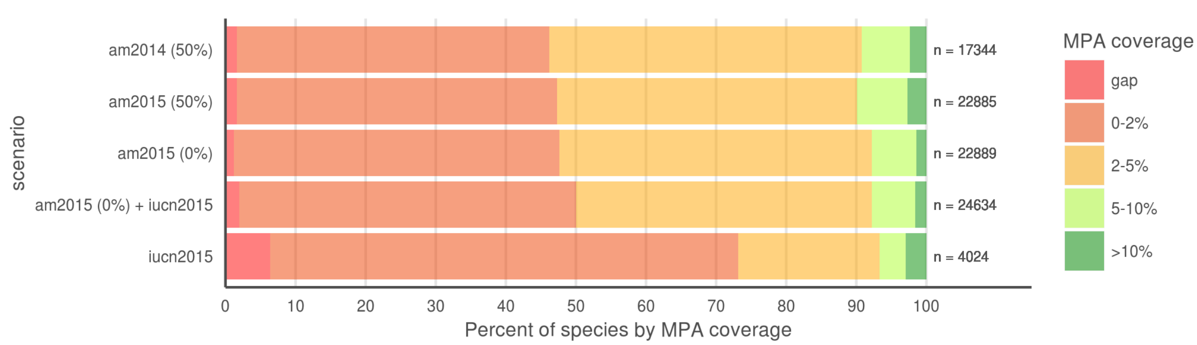


Fig. 4. Percent of species range covered by MPAs based upon methods in Klein et al. (2015). Scenario 1 replicates the original results, measuring protected range of species in AquaMaps 2014 dataset, with a 50% presence threshold, against the 2014 World Database of Protected Areas, filtered for IUCN categories I-IV that overlap marine areas. Scenario 2 updates the results using the 2015 AquaMaps dataset; the small changes despite the inclusion of an additional 5545 species demonstrates the robustness of the conclusions of Klein et al. (2015). Scenario 3, using 2015 AquaMaps data, drops the presence threshold to zero, showing an expected decrease in gap species, but also a decrease in species with 5% or greater protected range. Scenario 4 adds an additional 1745 species unique to the IUCN dataset, resulting in increases in gap species and species with less than 2% coverage. Scenario 5 examines species MPA coverage using only the IUCN dataset.

Scenario	n	gap	covered 0-2%	covered 2-5%	covered 5-10%	covered >10%
am2014 (50%)	17344	1.6	44.6	44.6	6.8	2.4
am2015 (50%)	22885	1.7	45.7	42.7	7.2	2.7
am2015 (0%)	22889	1.2	46.5	44.5	6.4	1.5
am2015 (0%) + iucn2015	24634	2.0	48.0	42.2	6.2	1.6
iucn2015	4024	6.4	66.8	20.1	3.8	2.9

7 References

still need to check citations, format properly, put into order

Kaschner K et al. (2015) AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication, www.aquamaps.org, Version 08/2015.

IUCN (2015). The IUCN Red List of Threatened Species. Version 2015-4. http://www.iucnredlist.org. Downloaded on 21 December 2015.

Halpern BS et al. "An index to assess the health and benefits of the global ocean." Nature 488.7413 (2012): 615-620. _doi:10.1038/nature11397._

Halpern BS et al. (2015). Patterns and emerging trends in global ocean health. PloS one, 10(3), e0117863.

Klein CJ et al. (2015). Shortfalls in the global protected area network at representing marine biodiversity. Scientific Reports 5: 17539. _doi:10.1038/srep17539._

Selig ER et al. (2013). Assessing global marine biodiversity status within a coupled socio-ecological perspective. PloS one, 8(4), e60284. _doi:10.1371/journal.pone.0060284._

Coll M et al. (2010). The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS ONE 5(8): e11842. *used AquaMaps to predict Med biodiversity. Also: Threshold = 0.*

*Martin CS et al. (2014). Manual of marine and coastal datasets of biodiversity importance. May 2014 release. Cambridge (UK): UNEP World Conservation Monitoring Centre. 28 pp. (+ 4 annexes totalling 174 pp. and one e-supplement). *review of marine data sets and data gaps etc, incl both IUCN and AM as well as many others*

Pimm SL et al. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. Science, 344(6187), 1246752. *uses range maps to show biodiversity areas; may use IUCN range maps. Also discusses gaps and possible things that can be done about them.*

Kaschner K, Tittensor DP, Ready J, Gerrodette T, Worm B (2011). Current and Future Patterns of Global Marine Mammal Biodiversity. PLoS ONE 6(5): e19653. *just what the title says - AM development, presence threshold 60%, also analyzes richness as a function of threshold*

Molinos JG et al. (2015). Climate velocity and the future global redistribution of marine biodiversity. Nature Climate Change. _doi:10.1038/nclimate2769._

Rondinini C, Wilson KA, Boitani L, Grantham H, & Possingham HP (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology letters, 9(10), 1136-1145. *compares point locality, range maps, and distribution models in terms of omission and commission errors; also outlines Extent of Occurrence and Area of Occupancy distinctions.*

Jetz W, Sekercioglu CH, and Watson JE (2008). Ecological correlates and conservation implications of overestimating species geographic ranges. Conservation Biology, 22(1), 110-119. *EOO maps are usually highly interpolated and overestimate small-scale occurrence, which may bias research outcomes*

Kaschner K, Watson R, Trites AW, Pauly D (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Marine Ecology Progress Series 316: 285–310. *check this citation journal name... This outlines the basic RES methodology - AM development*

Ready J et al. (2010). Predicting the distributions of marine organisms at the global scale. Ecological Modelling 221(3): 467-478. *Presents AM; assessing AquaMaps against other presence-only species models*

Spalding MD et al. (2007). Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. BioScience, 57(7), 573-583.

7.0.1 NOT CURRENTLY USED:

Jones MC, Dyeb SR, Pinnegar JK, Cheung WWL (2012). Modelling commercial fish distributions: Prediction and assessment using different approaches. Ecological Modelling 225(2012): 133-145. PDF *comparison of species distribution models including AquaMaps, Maxent and the Sea Around Us Project*

Hurlbert AH, and Jetz W (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proceedings of the National Academy of Sciences, 104(33), 13384-13389. _mostly rasters of range maps? "The scale dependence of range-map accuracy poses clear limitations on braod-scale ecological analyses and conservation assessments. ... we provide guidance about the approriate scale of their use_

---

8 Supplemental Information

9 Outline

9.1 Data processing

All code is publicly available and freely downloadable here on GitHub.

9.1.1 AquaMaps

AquaMaps data was provided by AquaMaps directly via ftp. Three SQL files were sent that contained (1) list of all 22889 species and their taxonomic information, (2) information for all (292,000?) half degree cells used for range mapping and (3) the probability of occurrence per cell for all 22889 species. These were extracted into .csv form using R Statistical Software.

Probably just cut, but maybe include in Suppl Materials. - BH: The release of AquaMaps distribution maps is not limited to comprehensively-assessed taxa, and maps are available across a much larger range of taxonomic classes; however, there is no guarantee that the list of species included within each class is a representative cross section of the entire class.

9.1.1.1 Possible figures

- Map of all cells
- Map of # of species per cell
- Example aquamaps (one species range)

9.1.2 IUCN

IUCN: While the polygons roughly define regions of presence/absence, additional attributes provide information on extant/extinct ranges, native/introduced ranges, and seasonality.

As of December 2015, IUCN had published species distribution maps for 8691 marine species across 24 taxonomic groups. For this analysis, we did not consider IUCN range maps for bird species, as those data are hosted separately by BirdLife International.

- which data sets are included?
- raster::extract() to convert polys to csvs
- which columns are included?

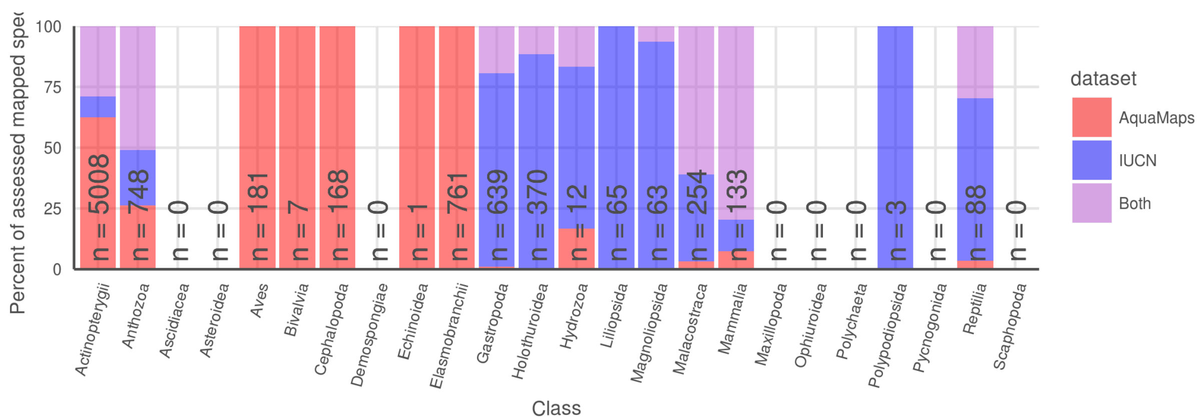
JA: I think here we only need a few sentences describing how we downloaded the data and on what date. Link to the IUCN spatial data website.

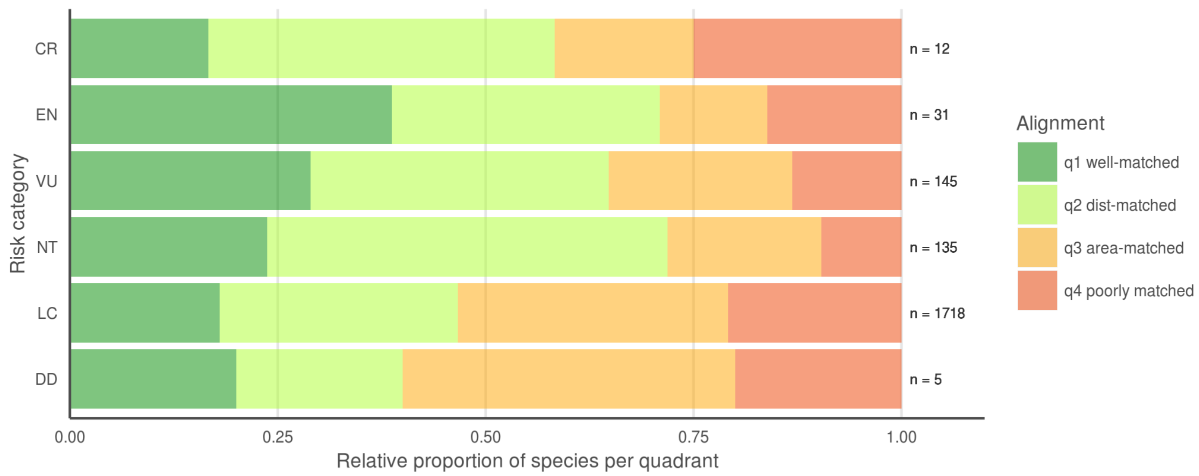
IUCN data is provided as polygon shapefiles. These vectors were rasterized to the same cell resolution as AquaMaps data (0.5 degrees).

9.1.2.1 Possible figures

9.1.2.2 Caveats For example, as of this writing, IUCN has released no spatial data for class Elasmobranchii (cartilaginous fishes including sharks and rays); and while IUCN offers a large number of maps within class Actinopterygii (ray-finned bony fishes), the available maps include only a few primarily tropical taxonomic sub-groups, such as wrasses, damselfish, butterflyfish, tunas, and billfishes, but are missing economically important subgroups including salmon, rockfish, and clupeids. However, IUCN's criterion of comprehensive assessment greatly reduces the risk of sample bias within the bounds of the assessed taxonomic groups.

Red List inclusion:



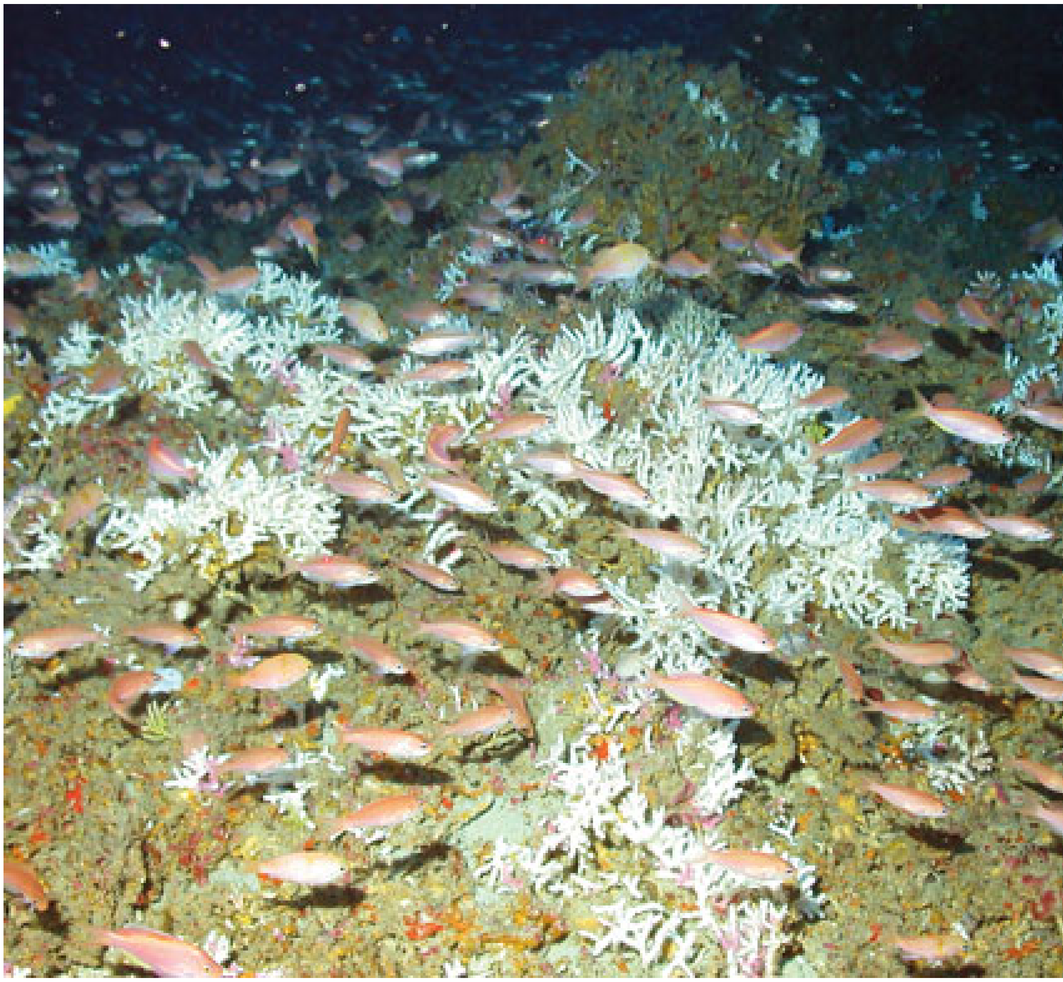


- Breaking down the quadrants by IUCN extinction risk categories (FIG 3c), we found that species with higher extinction risk tend to be better aligned between the two datasets, perhaps correlated to increased expert scrutiny. Does higher perceived risk lead to increased attention, and thus better understanding of species distribution? Or conversely, does increased attention to species distribution reveal more species at risk? Likely both mechanisms are at play on a case-by-case basis, depending on the species' taxon and region. *does this argument bear up to closer scrutiny? CR isn't dominated by Q1 any more*

table to show data poor status and reviewed status for AquaMaps maps represented in the quadrant plot - *perhaps update the quadrant plot to reveal data-poor species (and reviewed species? little overlap of reviewed & data-poor) instead of, say, red-list category which doesn't get discussed in the body of the paper? then this table can go in SOM if we like it*

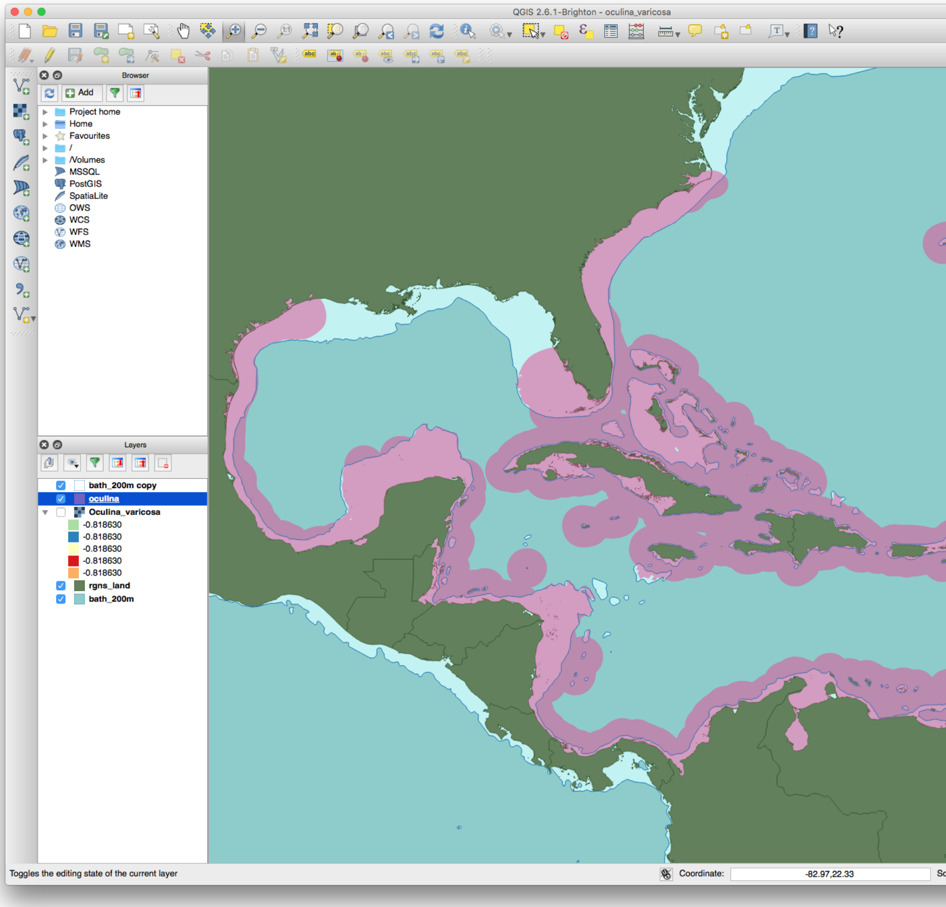
quadrant	n species	n data poor species	mean (median) data points	n reviewed species
all AM	22889	8749 (38.2%)	57.1 (16)	1296 (5.7%)
AM&IUCN	2166	457 (21.1%)	89.9 (33)	290 (13.4%)
q1	401	33 (8.2%)	233.0 (78)	100 (24.9%)
q2	682	151 (22.1%)	77.4 (39)	100 (14.7%)
q3	682	114 (16.5%)	52.4 (29)	65 (9.5%)
q4	410	159 (39.7%)	32.1 (13)	25 (6.2%)

9.2 illustrative maps for different quadrants and different mechanistic problems



Oculina varicosa - Ivory tree coral

- from IUCN: Colonies are found to depths of 152 m depth on limestone rubble, low-relief limestone outcrops, high-relief, steeply sloping prominences, and soft-bottom sloping habitats. Colonies are semi-isolated, patchy and low-growing in shallow water, or they form larger, massive coalescing aggregates (thickets or coppices) with substantial topographic relief in 50-100 m depth. In shallow waters (2-30m) the form is zooxanthellate, inhabiting limestone ledges. In deeper waters, an azooxanthellate form is known from the shelf edge off eastern Florida, USA from Ft. Pierce to Daytona (Reed 1980, 1983, 2002; Brooke and Young 2003).



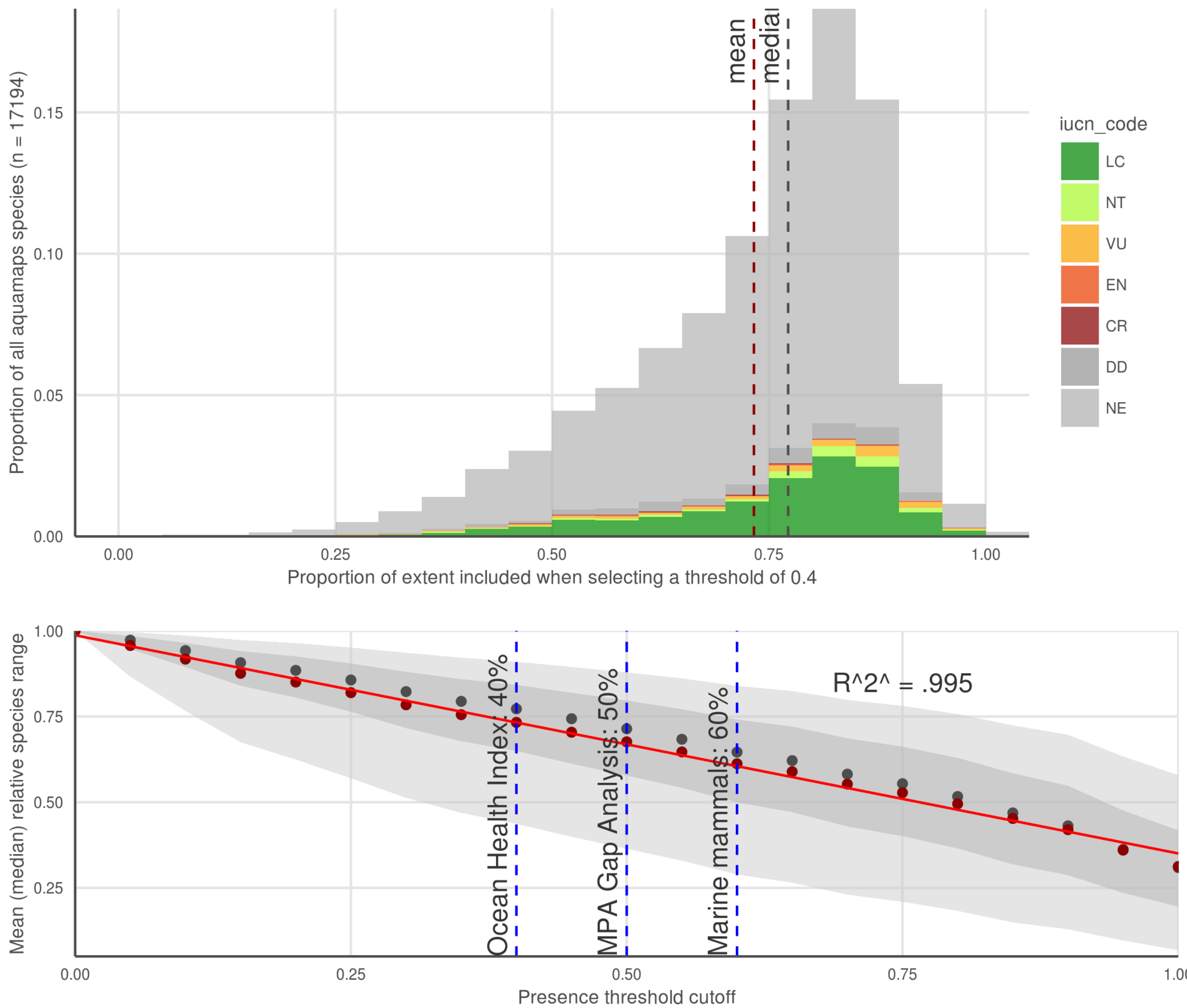
plotted against a 200 m bathymetry line:

9.3 AquaMaps: Effect of changing "presence" threshold on apparent distribution

9.3.1 AquaMaps presence threshold analysis - move to SOM

For our comparisons of global distribution of represented biodiversity and spatial alignment between datasets, we considered "present" to be any cell with a non-zero probability of occurrence, to best approximate the "extent of occurrence" as generally indicated by IUCN maps. To examine the effect of different presence threshold selections on the represented range of a species, we varied the threshold from 0.05 to 1.00 and calculated the average species range relative to a zero threshold.

9.3.2 Figure 5 (a, b):



AquaMaps distribution map extent remaining after applying a presence threshold. (a) A 40% threshold applied to all species in the AquaMaps dataset shows a mean loss of XXX, with a wide distribution in which some species lose nearly all of their apparent range. (b) Mean (median) remaining extent at increments of presence threshold. Dark grey ribbon includes 25% to 75% quantiles, while light grey ribbon includes 9% to 91% quantiles.

AquaMaps distribution maps indicate "probability of occurrence" within each 0.5° cell, with values ranging from zero to one, rather than a simple present/absent value as indicated by IUCN maps. Many studies convert this AquaMaps probability to a simple presence value by assigning a threshold value (REF references here). A higher threshold constrains an analysis to cells with near certainty of occurrence, while a low threshold captures larger areas of increasingly marginal suitability.

At a presence threshold of 40%, as used in the Ocean Health Index Species subgoal, the bulk of AquaMaps species suffer a significant decrease in represented range, and some species lose nearly their entire range. Incrementing the presence threshold from 0.00 to 1.00 for the entire AquaMaps dataset, the shallow downward trend indicates a low but consistent sensitivity to threshold choice, with no surprising tradeoffs that could suggest an "optimal" threshold. This pattern may not hold true for all subsets of AquaMaps species, however, whether subsetting by taxa or by georegion.