Aligning marine species range data to better serve science and conservation

Casey O'Hara[1], Jamie C. Afflerbach[1], Courtney Scarborough[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: 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 studies and conservation management decisions, yet most species ranges remain unknown, and existing range maps often suffer from data limitations and inconsistencies. AquaMaps and the International Union for Conservation of Nature (IUCN) are two distinct efforts to map marine species distributions at a global scale. Together these databases represent 24,637 species (92.9% within AquaMaps, 16.3% within IUCN), with only 2,279 shared species. Here we examine differences in predicted species ranges between the two datasets and find that these misalignments mainly result from divergent methodologies that introduce differing frequencies of commission and omission errors. We illustrate the scientific and management implications of these differences by repeating two recent applications - an assessment of global biodiversity within the Ocean Health Index and a global analysis of gaps in coverage of marine protected areas - and find significantly different results depending on how the two datasets were used. Until a single, highly accurate dataset of global marine species ranges becomes available, understanding the implications of dataset differences for conservation planning and decision-making remains essential.

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.*

Successful conservation of marine species critically depends on detailed and accurate understanding of where species exist and where they do not. By systematically comparing two widely used global datasets of marine species range maps, we find substantial differences that highlight inconsistencies in our understanding and communication of species ranges. Knowing where and why these inconsistencies occur will substantially improve our ability to develop effective and efficient conservation strategies that support resilient ocean ecosystems.

---

1 Introduction

Knowing where species exist and thrive 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) (IUCN 2015). These two spatial datasets have been used in hundreds of studies and applications for a wide range of purposes, including assessing marine species status (Halpern et al. 2012, 2015; Selig et al. 2013), evaluating global biodiversity patterns (Coll et al. 2010; Martin et al. 2014, Pimm et al. 2014, Kaschner et al. 2011), predicting range shifts (Molinos et al. 2015), and setting conservation priorities (Klein et al. 2015).

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 et al. 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; 92.9% 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 marine species). For the 2,279 species (9.3% 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 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 based on expert input of spatial boundaries of a given species' "limits of distribution" (IUCN 2015) - essentially a refined extent of occurrence, based on observation records and informed by expert understanding of species' range and habitat preferences. In contrast, AquaMaps models species distribution based on 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). Studies frequently define "presence" for AquaMaps by applying a probability threshold, e.g., the Ocean Health Index defines presence as 40% or greater probability of occurrence (Halpern et al. 2012, 2015).

The methodologies behind the creation of these datasets imply differences in prediction of species distribution due to errors of commission (falsely indicating species presence) and omission (falsely indicating species absence). Geographic range data such as IUCN range maps frequently introduce commission errors, while species distribution models such as AquaMaps will likely introduce 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). By comparing maps resulting from IUCN and AquaMaps methodolog, we can identify and possibly address mechanistic causes for each type of error.

The two datasets have notably different taxonomic (Fig. 1) and regional (Fig. 2) coverage. AquaMaps encompasses a broader range of taxa than IUCN, as IUCN spatial data files are only available for select taxonomic groups that have been comprehensively assessed. While species numbers in both datasets peak in tropical latitudes near the equator, species counts for IUCN maps drop quickly beyond 30°N and 30°S, while species counts for AquaMaps remain 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. To achieve more comprehensive global coverage of species ranges these two datasets can be used together, 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 (how much of the smaller range falls within the larger range, i.e., where on the map) against the area alignment (ratio of smaller range area to larger range area, i.e., how much of the map) for each shared species (Fig. 3A). This analysis revealed a general negative linear pattern, suggesting that increasing similarity in range area correlates with decreasing distribution alignment. AquaMaps tends to extrapolate species ranges into suitable areas beyond known occurrences, such that each additional unit of range predicted by AquaMaps will fall in different locations than an additional unit of range predicted using IUCN methodology. For species with dissimilar range areas, predicted distribution for the smaller range can more easily fall within the generous bounds of the larger range. For species with increasingly similar range areas, differences in methodology become more difficult to "hide," and the distribution alignment generally becomes poorer.

The mean distribution alignment for species included in both datasets was 63%; the mean area alignment was 54.5%. By dividing the map-paired species into quadrants based on these means, we highlight categories of relationships that help further explain this general pattern. For representative maps from each category, please refer to the supplemental materials (Fig. S1).

The upper right quadrant includes species (n = 527) whose described ranges are above average 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. 3B). This result is not surprising, as species with very large ranges are likely to be more aligned regardless of methodology simply because their ranges span nearly the entire map. We expected to see significant concurrence between maps for IUCN Red List species of concern, assuming that increased attention would lead to improved understanding of species ranges; however, we found no compelling patterns to indicate that this was the case (Fig. S2).

The area-mismatched ranges contained in the upper left quadrant (n = 709) 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 = 237; 33.4% 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 (Fig. S3). 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 = 635) 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 areas are meaningless when the distributions are poorly aligned.

The most vexing cases are in the lower left quadrant (n = 443), where neither distribution nor area match well. Data-poor species are more common in this quadrant; indeed, the median number of species occurrence records (averaging OBIS and GBIF occurrences) for this quadrant is 24 records, compared to a median of 97 records 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

Method-driven differences in commission and omission errors drive clear and significant differences in species range descriptions between AquaMaps and IUCN datasets. 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 Red List-assessed species in both IUCN and AquaMaps datasets (n = 7,963), prioritizing IUCN data for the 2,026 species included in both sources. OHI currently uses a probability of occurrence threshold of 40% to determine species presence for AquaMaps data.

We calculated 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). Reducing the threshold increases the apparent range of a species; the slight decrease in average score for scenario 1 suggests increased spatial representation of threatened species. The slight increase in mean score for scenario 2 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. The large decrease shown in scenario 3 indicates that a zero threshold greatly increases the spatial representation of high-risk species.

Given that only 25.4% of species overlap, it is surprising that changing the priority for overlapping species from IUCN maps to AquaMaps would result in such large country-level score shifts as seen in scenarios 2 and 3. While the mean global score did not vary significantly from scenario to scenario, select countries gained up to 7.5 points while others dropped as many as 5 points. This result indicates that especially on a national or regional scale, an arbitrary change in how the two datasets are combined can result in a very different assessment of species conservation status.

Case Study: MPA Gap Analysis

Klein et al. (2015) compare the global distribution of species to the global distribution of marine protected areas to assess how well the MPAs protect key species and identify which species fall through gaps in protection. The study relied on the AquaMaps database, using a probability of occurrence threshold of 50% or greater, to determine species presence, and the World Database of Protected Areas to define zones of marine protection. They found that the global MPA network leaves 90.5% of marine species with less than 5% of their overall range represented within MPAs, and 1.4% of species have no protection at all ("gap" species).

We recalculated the amount of under-protected and gap species using either IUCN or AquaMaps data (using 2015 AquaMaps data and a 0% threshold to allow the most meaningful comparison to IUCN's "limits of distribution", Fig. 4). We found a five-fold increase in the proportion of gap species (6.4% of species vs. 1.2%) and dramatically larger proportion of species with less than 2% of their range protected (73.2% of species vs. 47.7%). However, this comparison also indicates a larger proportion of well-protected species with greater than 10% of range protected (2.9% of species 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 show reasonable agreement for many well-studied species, but substantial differences illustrate uncertainty in our understanding of spatial distribution for many others. Although many other appriaches exist for species distribution modeling, these two are the only ones applied globally to marine species. Identifying and addressing differences in these datasets will increase their utility for research and conservation actions. Several likely drivers of commission errors and omission errors between these datasets point to a few important ways to improve range alignment.

For IUCN range data, clipping ranges to known depth limits improves output for many species, most notably corals and reef-associated fishes. If species' depth limits are not known, simple rules of thumb will likely reduce commission errors without introducing substantial omission errors. For example, for most corals, researchers could clip range maps to the photosynthetic limit of 200 meters. For AquaMaps range data, dependent primarily on environmental and physical preferences and conditions, implementing area restrictions based on biogeographical criteria such as Marine Ecoregions of the World (Spalding et al. 2007) would likely decrease commission errors and improve predictive power, especially for data-poor species.

Many studies using AquaMaps data test the sensitivity of results by varying the probability of occurrence threshold to determine presence. This decision ultimately represents a tradeoff between errors of commission (low threshold) and omission (high threshold). Using AquaMaps in conjunction with IUCN can mitigate potential errors, while also increasing the taxonomic and spatial breadth of coverage, as long as the differences between the datasets can be reasonably minimized. In this case, we recommend a presence threshold of 0% as it most closely approximates the "limits of distribution" criterion defined by IUCN data providers.

Effective management and protection of marine species depends on a robust understanding of where species exist and where they do not; without this knowledge we risk wasting resources protecting low-value regions while missing opportunities to protect critical ones. By identifying the differences between these two fundamental marine species range datasets and understanding the likely mechanisms causing these discrepancies, we improve our ability to develop strategic and effective conservation policy that supports a resilient ocean ecosystem.

---

5 Methods

Comparison of taxonomic and regional distribution: To examine the overall taxonomic distribution across the spatial datasets (Fig. 1), 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 number of species per cell was simply the count of the species present. For the AquaMaps dataset, we determined per-cell species count by including 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 count for each dataset by plotting average species count against latitude and longitude (Fig. 2).

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. 3A:

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

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

For each species with ranges described in both IUCN and AquaMaps, A_{small} and A_{large} indicate the smaller and larger range representation (regardless of which dataset). $A_{small \cap large}$ represents the amount of overlapping area between the two datasets. 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 (which prioritizes 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 2,279 overlapping species and again compared against the protected area raster.

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

---

6 Figures and captions

6.0.1 Fig. 1

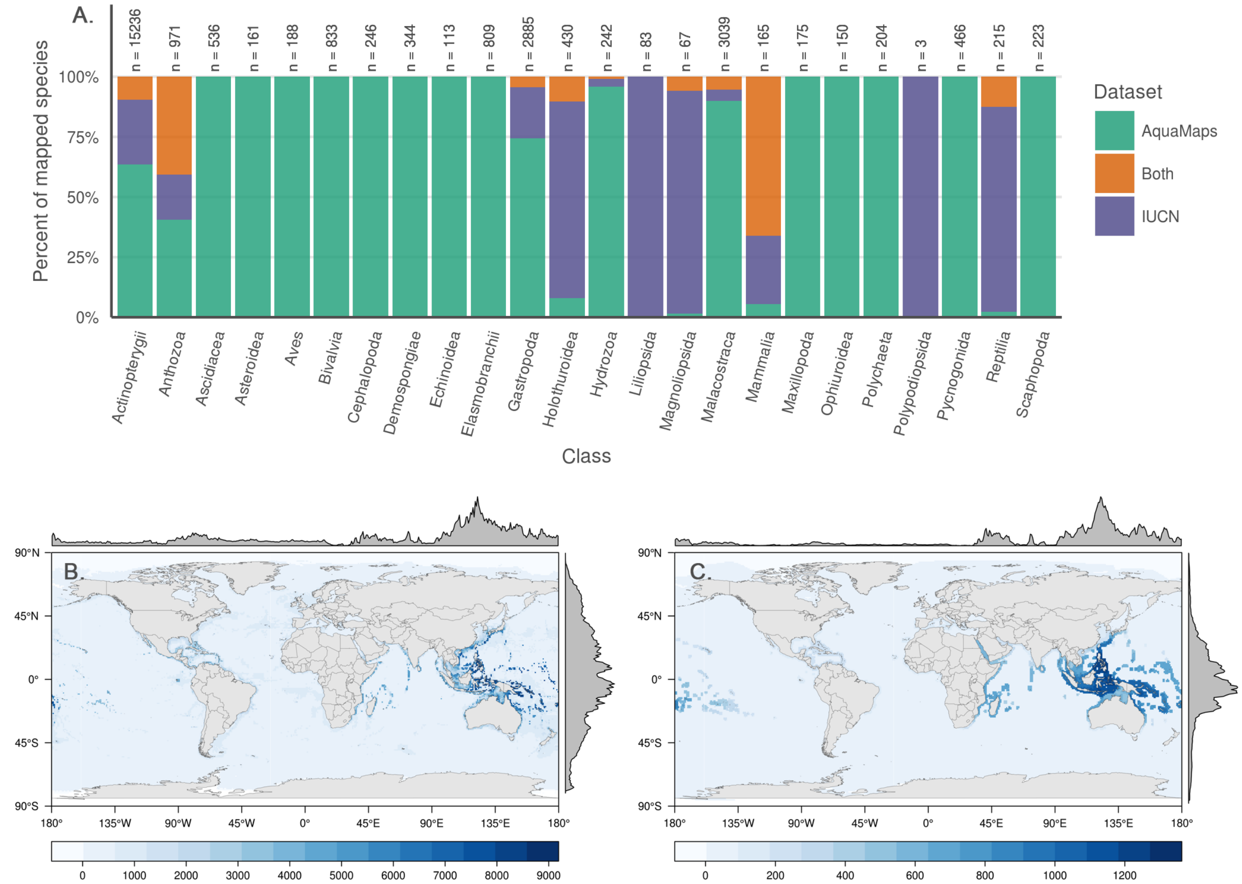


Fig. 1. (A) Number and proportion of species by taxa included in each dataset. Overlapping species are dominated by bony fishes (983 species, primarily tropical taxa) and corals (396 species). (B, C) Number of global marine species according to (A) AquaMaps and (B) IUCN. The margin frequency plots show relative species count per cell at each latitude and longitude.

6.0.2 Fig. 2

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 2,279 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 = 466; 20.1 %). The upper left quadrant (quadrant 2) comprises species whose maps agree well in distribution, but disagree in extent (n = 687; 29.7 %). 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 = 691; 29.9 %). The lower left quadrant (quadrant 4) indicates species for which the map pairs agree poorly in both area and distribution (n = 470; 20.3 %).
(B) Alignment quadrant breakdown of paired-map species by taxonomic group.

6.0.3 Figure 3

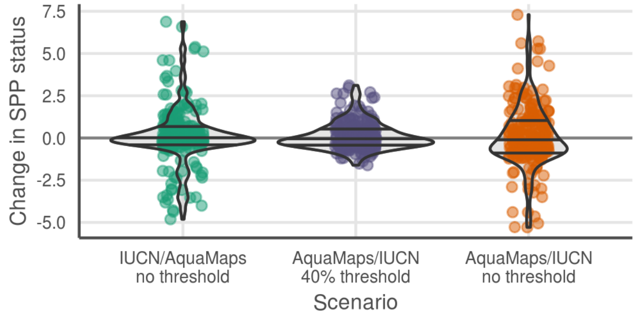


Fig. 3. Change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios. Scenario 1 shows the effect of reducing the probability threshold to 0% for AquaMaps presence to more accurately track the definition of IUCN "limits of distribution." Scenario 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.
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.

6.0.4 Figure 4

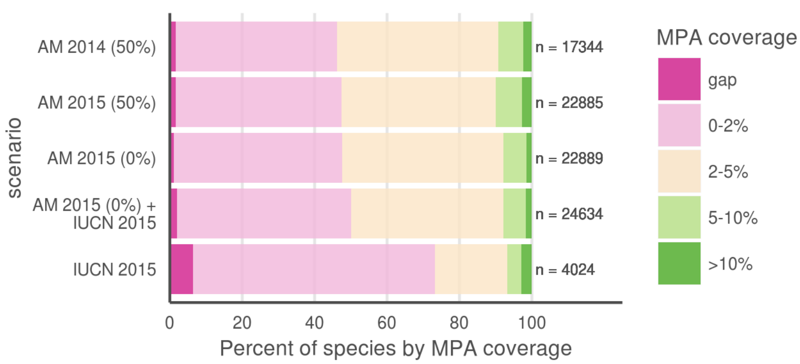


Fig. 4. Percent of species range covered by MPAs based upon methods in Klein et al. (2015). Scenario 1 replicates the original results, measuring protected range of species in AquaMaps 2014 dataset, with a 50% presence threshold, against the 2014 World Database of Protected Areas, filtered for IUCN categories I-IV that overlap marine areas. Scenario 2 updates the results using AquaMaps 2015, showing very small changes despite the inclusion of an additional 5,545 species. 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 IUCN, resulting in increases in gap species and species with less than 2% coverage. Scenario 5 examines species MPA coverage using only the IUCN dataset.

7 Acknowledgments

8 References

double check formats; also - include DOIs for references? which if any?

Kaschner K et al. (2015) AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication, www.aquamaps.org, Version 08/2015. ***since this is not a peer-reviewed publication, can we include this for PNAS reference list?***

IUCN (2015). The IUCN Red List of Threatened Species. Version 2015-4. http://www.iucnredlist.org. Downloaded on 21 December 2015. ***since this is not a peer-reviewed publication, can we include this for PNAS reference list?***

Halpern BS et al. (2012). An index to assess the health and benefits of the global ocean. Nature, 488(7413), 615-620. _doi:10.1038/nature11397._

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

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.

8.0.5 Datasets and tools: cite these?

- WDPA IUCN and UNEP-WCMC (2014). The World Database on Protected Areas (WDPA) [On-line]. Cambridge, UK: UNEP- WCMC. Available at: www.protectedplanet.net [Accessed December 1, 2014].
- R R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- QGIS QGIS Development Team (2015). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

8.0.6 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_

---

9 Supplemental Information

10 Outline

10.1 Data processing - should this go in methods???

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

10.1.1 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 (259,200?) 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.

10.1.2 IUCN

As of December 2015, IUCN had published species distribution maps for 4027 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.

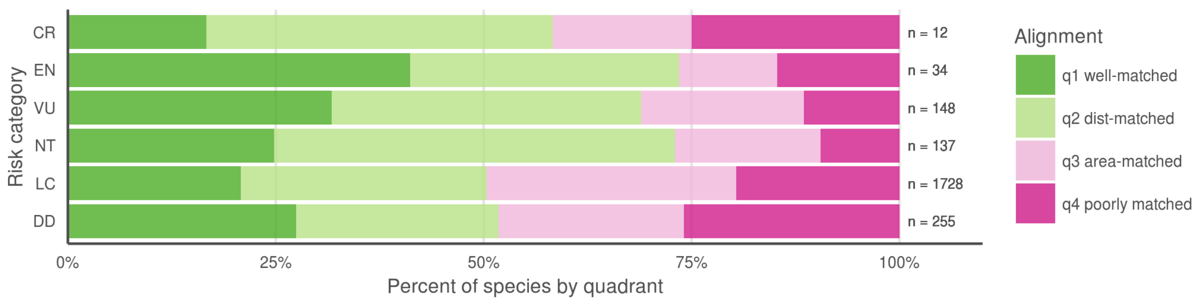
Each shapefile contains polygon limits of distribution indicating regions of presence/absence, additional attributes provide information on extant/extinct ranges, native/introduced ranges, and seasonality.

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

10.2 S1: sample maps from each quadrant

a) quad 1: Kajikia albida
b) quad 2: Conus episcopatus
c) quad 3: Conus magnificus
d) quad 4: Acanthurus nigroris

10.3 S2: risk by quadrant

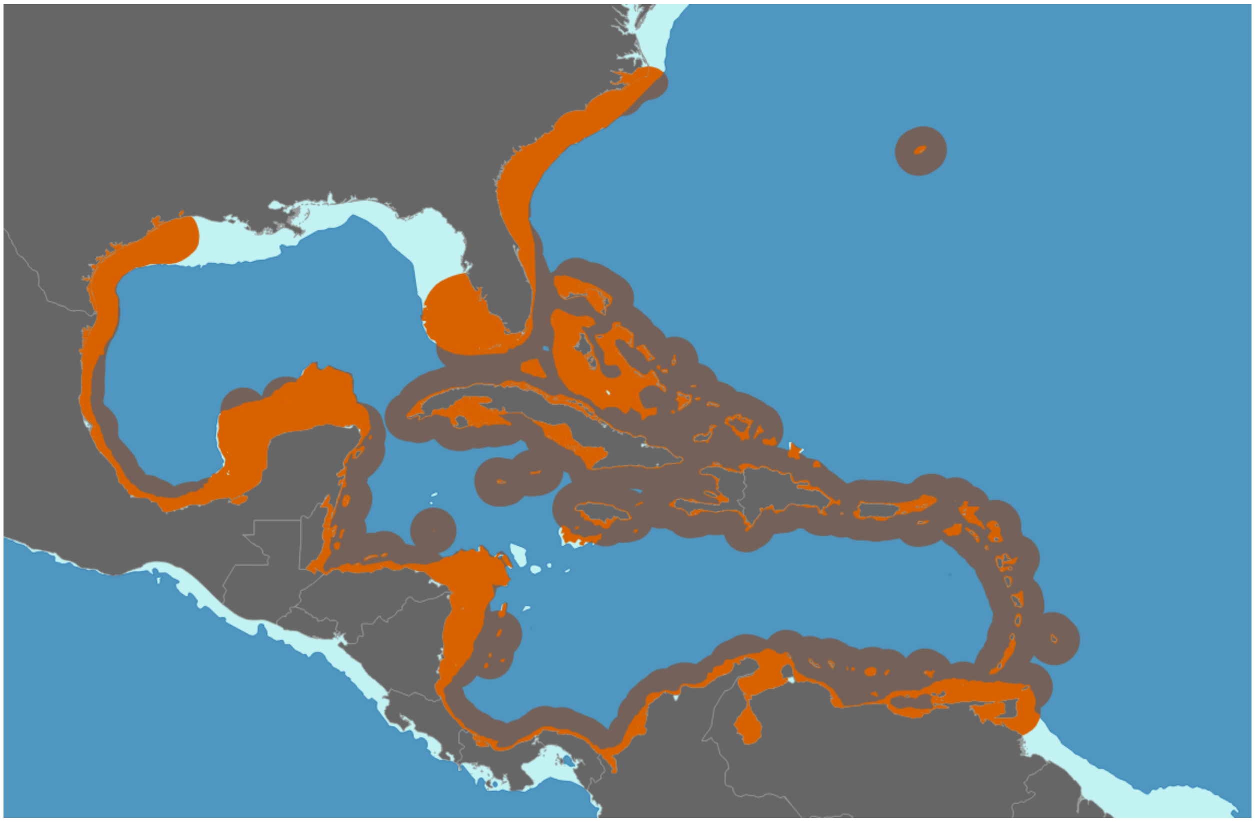


- Breaking down the quadrants by IUCN extinction risk categories (Fig. S2), we found little support for our hypothesis that maps for species with higher extinction risk tend to be better aligned between the two datasets, perhaps correlated to increased expert scrutiny. Does higher perceived risk lead 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*

10.4 S3: coral depth map

- 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).

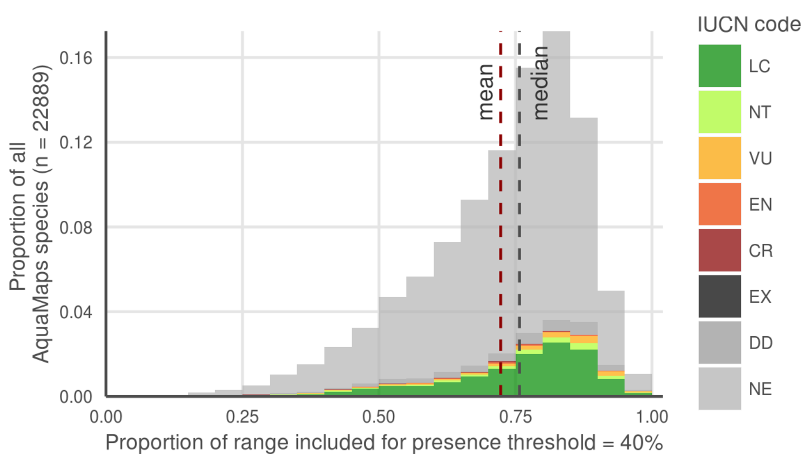
plotted against a 200 m bathymetry line:



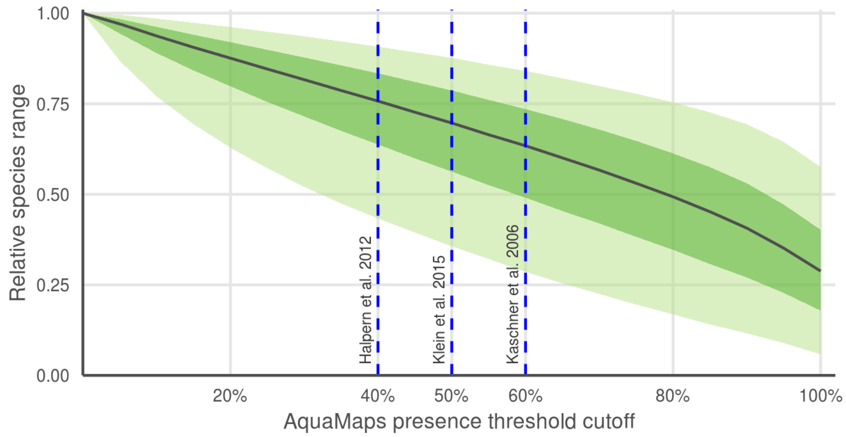
10.5 S4: AquaMaps threshold examination

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 "limits of distribution" as 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% to 100% and calculated the average species range relative to a zero threshold.

S4a:



S4b:



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 5% to 95% 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.

10.6 MPA gap analysis - explanation

why so many gaps and 10+ species show up with IUCN. I will write up a paragraph once I've had a little more time to examine maps

- check which species are gap species; check maps
- check which species are 10%+ species and check maps
- where do these species fall georegionally? is there anything odd about their ranges? any reason why IUCN would be
- hypothesis to explain more gap species: Aquamaps more diffuse than IUCN, catching individual cells that could be suitable; a single cell (even low probability) moves a species from "gap" to non-gap status. IUCN more clustered by polygon boundaries; so no scattered cells to accidentally fall into an MPA.
    - introducing errors: AM = likely commission (overestimates species with tiny bits of range in MPAs) vs IUCN = possible omission (misses possible real habitat that falls within MPAs)
- hypothesis to explain more 10%+ species - ten random samples most have IUCN polygons just east of Australia where a large MPA is indicated - Great Barrier Reef I assume - a single large poly falling in GBR gets a high score, while for the same species, many of the AM ranges (also in same area) scatter suitable area over larger extent, so lower percent of species range is protected

- introducing errors: IUCN = likely omission (skipping possible real habitat outside of MPAs) and likely commission (possibly overcounting real habitat in MPAs e.g. GBR); AquaMaps = probably a little commission (possibly underestimating, if it overpredicts lots of range outside MPAs, thus reducing the proportion of protected range)