Aligning species range data to better serve science and conservation

Casey O'Hara, Jamie Afflerbach, Ben Halpern

0.1 Abstract

questions for abstract style: voice (passive vs active, 'we' vs 'this paper'), tense (present vs past)...

Two global datasets of species distribution maps - AquaMaps and the International Union for Conservation of Nature - dominate our understanding of marine species ranges throughout the world's oceans, and inform a wide range of biodiversity and conservation purposes. This paper examines differences in predicted species ranges between the two datasets, proposes mechanistic causes and potential solutions for such differences, and explores the implications of these differences for management and conservation decisions.

Together, the two datasets provide range information for 24520 unique species, with only 2166 species common to both datasets. When selecting one dataset over the other, spatial and taxonomic coverage should be considered against the scope of the project.

To understand mechanistic reasons for differences in species range predictions, we compare the distribution and extent of the pair of predicted ranges for the subset of species included in both datasets. Categorizing the results based on how well each pair of species maps aligns, we identify several issues that commonly lead to misalignment in species range predictions. Notably, IUCN maps often disregard bathymetry for depth-limited species, leading to predictions of species presence at unsuitable depths, and AquaMaps maps for data-poor species often extrapolate presence far afield from known occurrences. Understanding these issues points toward solutions for providers and users of these datasets.

We reexamine two recent global analyses that depend on these datasets: the Ocean Health Index biodiversity goal and a global analysis of gaps in coverage of marine protected areas. Variations in which dataset is used, and how, lead to significantly different estimates of species diversity and conservation effectiveness.

0.2 Significance

subset of the abstract

---

1 Introduction

Peer beneath the waves anywhere in the world - a moody kelp forest, a bustling coral reefscape, the frigid depths beneath a polar ice sheet, the endless blue of the open ocean - and you will find a unique ecosystem defined as much by which species are absent as by which species are present. Knowing where individuals of a species exist, and where they 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 modeled species distribution maps (Kaschner et al. 2013) and International Union for Conservation of Nature (IUCN) species range maps (REF). These two spatial datasets are used for a wide range of purposes, including assessing marine species status (Halpern et al. 2012, Selig et al. 2013), evaluating global biodiversity patterns (Coll et al. 2010, Martin et al. 2014), 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 2008?)

To understand the implications of differences between these two datasets, we compared how each data set represents the global spatial and taxonomic distribution of species. For the relatively small number of species mapped in both datasets, we examined how well the species maps align and determined several issues that lead to misalignment between predicted species distributions. We then subjected two recent marine biodiversity studies - the Ocean Health Index biodiversity goal (Halpern et al. 2012) and a global analysis of gaps in protection afforded by marine protected areas (MPAs) (Klein et al. 2015) - to a sensitivity analysis, substituting one dataset over the other, to highlight the possible consequences of different data use decisions on our understanding of *the status of* marine biodiversity.

1.1 Overview of AquaMaps and IUCN datasets

The IUCN publishes species range maps as spatial vector polygon shapefiles, bundled by taxonomic groups. Using GIS, experts outline spatial polygons to represent a given species' extent of occurrence, based on observation records and refined by expert understanding of the species' range and habitat preferences. *IUCN releases a range map bundle for a taxonomic group once it has been "comprehensively assessed," i.e. at least 90% of the species within the taxonomic group have been evaluated (REF).* While this mitigates sampling bias within taxa, it also means that entire taxonomic groups remain unavailable until they have met this threshold of comprehensive assessment.

AquaMaps develops species distribution maps based on modeled relative environmental suitability. For each mapped species, environmental preferences (e.g. temperature, depth, salinity) are 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. A subset of the resulting distribution maps (1296 of 22889 species maps; 5.7%) have been checked or formally reviewed by experts, to validate and improve the parameters and underlying model.

IUCN range maps are intended to include all possible regions in which a species may be present, with the caveat that this does not imply the species is evenly distributed everywhere within the boundaries of the map(REF). AquaMaps modeled distribution maps provide a more nuanced prediction of species presence within the extent of occurrence, but the model may not capture all the complexities that drive species distribution (REF). Due to these differences in methodology and intent, IUCN range maps are more likely to overpredict species presence (and underpredict absence) while AquaMaps modeled distribution maps are relatively more likely to underpredict species presence.(REF and REF)

---

2 Results and Discussion

2.1 Comparing the two datasets

2.1.1 Taxonomic distribution between datasets

I wonder if this section is becoming less a part of the "story", maybe this part should go in SOM? *- though I think it's pretty important to understand that there is a huge difference in which species and taxonomic groups are covered in each set. That also helps to explain why there is such a small overlap, and helps show that the overlapping species are not necessarily representative? It also provides a logical lead-in from description of the overall datasets to the analysis of just the overlapping portion of the datasets.*

Each dataset offers spatial distribution information for large numbers of species (22889 AquaMaps-mapped species; 4138 IUCN-mapped marine species). However, the datasets vary in terms of taxonomic coverage and regional coverage, with only 2166 species included in both datasets (9.5% of AquaMaps species; 52.3% of IUCN marine species; Figure 1a).

The distribution of IUCN-mapped species skews toward tropical latitudes and away from the Atlantic and Eastern Pacific compared to the distribution of AquaMaps-mapped species.(Figure 1b and 1c) This skew likely reflects the fact that the IUCN dataset focuses more heavily on coral reef-associated taxa than does the AquaMaps dataset (see Figure 1a).

For spatial assessments of biodiversity, the choice of one dataset over the other is likely to result in significantly different conclusions. For studies confined to a narrow range of taxa or to a narrow spatial scale, one dataset may offer an advantage over the other in the number of mapped species available. For global scale biodiversity studies, however, the selection of one dataset over the other will entail tradeoffs in spatial coverage and taxonomic breadth versus depth.*use something from Rondinini here?*

Small overlap means that using both datasets in conjunction will add a huge number of species without duplicating efforts, if the differences between the two methods can be reconciled appropriately

find a reference that describes what makes a "good" dataset for global biodiversity, e.g. OHI, or maybe species richness vs diversity vs "health" or whatnot: (Tittensor et al., 2010)

2.1.2 Defining spatial alignment between the two datasets

The IUCN and AquaMaps spatial datasets share 2166 species in common. Placing each of these species' pair of maps side by side, we would hope to see spatial correlation both in the global pattern of species distribution (where on the map) and the extent of species range (how much of the map).

Dividing the map-paired species into quadrants based on median values for each dimension allows us to examine the *implications* of four different qualities of alignment. (Figure 2a)

Quadrant 1 highlights species whose map pairs agree in both spatial distribution and the extent of described ranges, as we would expect for well-understood species. Examining spatial alignment by taxonomic group (Figure 2b), we found that certain taxa were far more likely than others to be spatially well-aligned; in particular, wide-ranging pelagic organisms (50% or more of ocean area) such as marine mammals, tunas, and billfishes were more consistently well-aligned than demersal and reef organisms. *propose an explanation here? basically - these species go pretty much anywhere, so both maps predict most of the ocean, you'll get a good overlap - not really that revolutionary*

The extent-misaligned maps contained in quadrant 2 provide a richer set of examples to understand the fundamental differences between the datasets; examining these, we can identify several mechanisms that could cause such extent misalignment.

For 630 of 682 species in quadrant 2 (92.4%), the IUCN extent is larger than the AquaMaps extent. For many of these species, a quick look at the maps (see SOM for examples) shows that the IUCN range hews closely to the AquaMaps range, while adding a significant buffer zone. Coral species dominate this quadrant (294 coral species - 58.2% of all corals and 43.1% of all species in quadrant 2). Most corals and reef-associated organisms prefer shallower waters, and we suspect that many of these extent-misaligned map pairs can be easily explained: seafloor depth is explicitly included in AquaMaps models, while depth is frequently ignored as a factor in IUCN range maps, leading to overprediction of species range into areas too deep to support the species. (For examples and details, see SOM)

- ***verify some of the coral maps by saving raster, showing raster and polygons in QGIS against a bathymetry layer*** - put a map for Oculina varicosa (and others?) in the SOM
- ***include recommendation for IUCN range maps to be clipped to bathymetry where appropriate? or save til end?***

The extent-alignment of species maps found in quadrant 3 can in many cases be attributed to "two wrongs make a right." For these species, IUCN ranges frequently overextend into unsuitable habitat, as in the case of many quadrant 2 species, while at the same time, AquaMaps ranges often aggressively extrapolate presence into locations where IUCN predicts absence. This results in an area ratio close to 100%, giving the impression of extent alignment. ***this quadrant seems to be a false alignment then? should there be a stronger warning here***

Quadrant 4 species maps suffer from the same issues as those in quadrant 3, but here the AquaMaps extrapolation extends even further, pushing the extent alignment ratio lower. In this quadrant, AquaMaps predicts a larger extent than IUCN for 70% of species represented in this quadrant, compared with just 30% of species in the entire set of overlapping maps *(this is supposed to show that much of this quadrant is based on overextrapolation of AquaMaps... not sure that's clear yet)*. AquaMaps range maps in quadrant 4 include a high proportion of data-poor species in which the environmental envelope model is based on fewer than 10 known cell occurrences, and at the same time include fewer reviewed species maps than the other quadrants.

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	n reviewed species
overall	22889	8749 (38.2%)	1296 (5.7%)
AM&IUCN	2166	457 (21.1%)	290 (13.4%)
q1	401	33 (8.2%)	100 (24.9%)
q2	682	151 (22.1%)	100 (14.7%)
q3	682	114 (16.5%)	65 (9.5%)
q4	410	159 (39.7%)	25 (6.2%)

data poor = fewer than 10 points used to generate the environmental envelopes; reviewed = map checked or formally expert-reviewed

Other hypotheses/case studies to highlight other possible mechanistic differences? I think we found a couple of the big ones; good enough?

2.2 Application to OHI

The global Ocean Health Index (OHI) (Halpern et al. 2012), a composite index made up of 10 goals, utilizes both of these datasets to inform the Species subgoal of the Biodiversity goal. As it is currently calculated, the Species subgoal uses species spatial distribution data and IUCN Red List conservation status to calculate an area-weighted mean species status in each of 221 exclusive economic zones. Spatial distributions were gleaned from both IUCN and AquaMaps datasets, preferring IUCN data for species represented in both data sets. OHI uses a probability threshold of 40% to determine species presence for AquaMaps data. Species with no spatial data in either dataset were excluded, as were species with insufficient information to determine conservation status (including species listed as not evaluated or data deficient).

Briefly summarize results

Since the Ocean Health Index Species subgoal relies on spatial data from both datasets, the impacts of these threshold and preference changes will be somewhat muted (Figure 3). When IUCN range maps are the preferred data source, only the subset of AquaMaps-only species will be affected by threshold changes; and when AquaMaps is the preferred source, the IUCN-only species will dampen the effect of a threshold change.

2.3 Application to 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 the gaps in protection. The MPA gap analysis relied on the AquaMaps database available in early 2015, comprising distribution maps for XXX species. Species presence was defined using a probability threshold of 50% or greater. The study defined four subsets of protected areas using the World Database of Protected Areas, based upon combinations of IUCN protected area classification, marine designation, and spatial marine coverage. The study determined that currently defined marine protected areas (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 marine protected areas.

working on the analysis…

- rerun based on old AquaMaps at 50% (to verify methodology gets a close result to paper), then new AquaMaps for updated baseline
  - consider also: filter on STATUS == 'Designated', and filter out those pesky fishery management areas and manatee speed zones…
- run with threshold at 0% (not all the same as original paper)
- run with filtering out data-poor species?
- run with IUCN maps instead of AquaMaps
- Which aspects of the analysis should we focus on?:
  - overall protection, for 4th subset of protected areas? or overall for all subsets? (see paper, table 1)
  - overall vs broken down by taxonomic groups? (see paper, figure 1)
  - spatial distribution of gap species (see paper, figure 2)

Predictions: What happens when using a 0% threshold instead of a 50%? What happens when using IUCN rather than AM? IUCN tends to overestimate extent esp for coastal species? More species will be included in MPAs so fewer apparent gap species. Included range area inside MPAs will increase. Policy implications?

2.4 Wrap it up

Recommendations on data use?

- IUCN mapmakers
  - at least clip polygons to bathymetry for depth-sensitive species, come on now people, it's really pretty obvious.

  - Note something about how photosynthetic marine species, and the species that depend on them, are far more sensitive to bathymetry than terrestrial species are to altitude...
  - Note that this is explicitly mentioned in the training for IUCN red list marine maps
    * "bathymetry can be used to delineate species' ranges limited by depth in the same way as elevation is used for terrestrial species"
- AquaMaps mapmakers
  - get expert eyeballs on all the maps
  - revisit the decision to use data-poor species... or at least make it more up-front about the ramifications of including the data-poor species
- Users
  - understand that range maps tend to overestimate presence; AquaMaps are less likely to overestimate presence, but trade off for overestimating absence. Both better than simple point locality data.
  - to use both together:
    * consider setting AM presence threshold to 0% to create an extent of occurrence map, more like IUCN map.
    * consider additional filters and criteria, e.g. `reviewed` and/or `occurcells` to avoid over-extrapolation problem
    * consider clipping IUCN ranges to bathymetry for depth-limited species or taxa
    * or consider de-weighting IUCN for area-weighted calcs, for taxa in which IUCN is consistently overestimating presence

---

3 Methods

3.0.1 taxonomic distribution comparison

To examine the overall taxonomic distribution across the spatial datasets, 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. We then filtered the species list to those that have been evaluated for the IUCN Red List of Threatened Species (in SOM).

Currently this is also included in body - take out and leave in here? Per BH: "Either cut or move to data description section of methods above.": The IUCN releases spatial data sets when a taxonomic group (typically on the scale of order or family) has been comprehensively assessed, to guard against sample bias (though non-comprehensive datasets are available for reptiles and marine fish). As such, spatial data for many taxonomic classes remain unavailable, and within a class, the assessed sub-groups may not represent the entire class.

3.0.2 global spatial distribution comparison

To compare the spatial representation of the two datasets directly, we first 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.

map pairs comparison

Using genus and species binomials to identify paired maps, we selected the subset of marine species that have range maps in both IUCN and AquaMaps current native distribution (n = 2166). We used the same criteria as outlined above to determine species presence within a cell.

from results/discussion: paraphrase or expand upon this Overlaying paired distribution maps for a given species, we calculated two dimensions of spatial alignment: *distribution alignment*, which we defined as the proportion of the smaller range intersecting the larger range; and *area alignment*, which we defined as the ratio of the smaller range to the larger range.

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

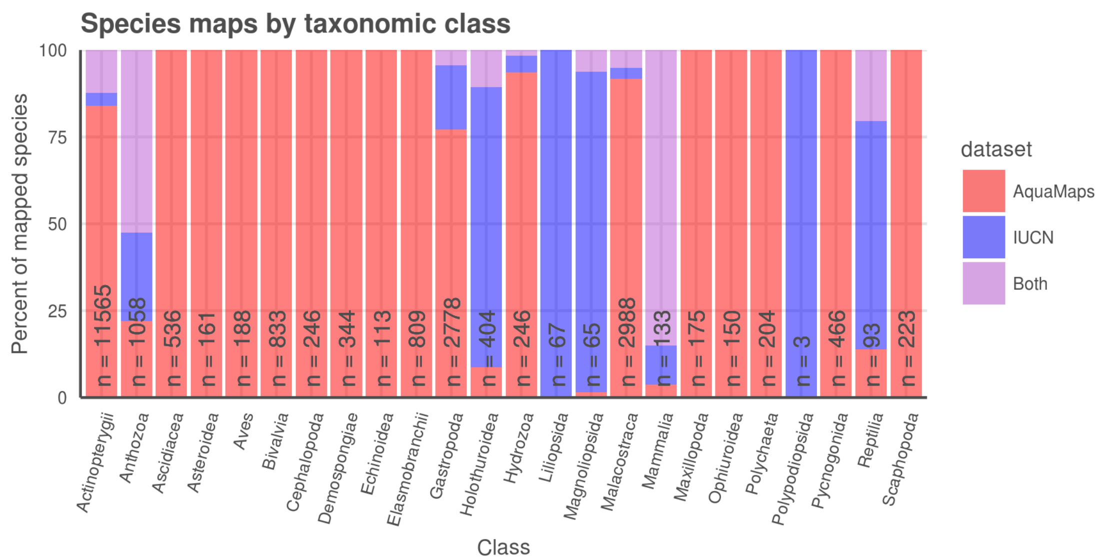
$$\alpha_{area} = \frac{A_{small}}{A_{large}} * 100\%$$

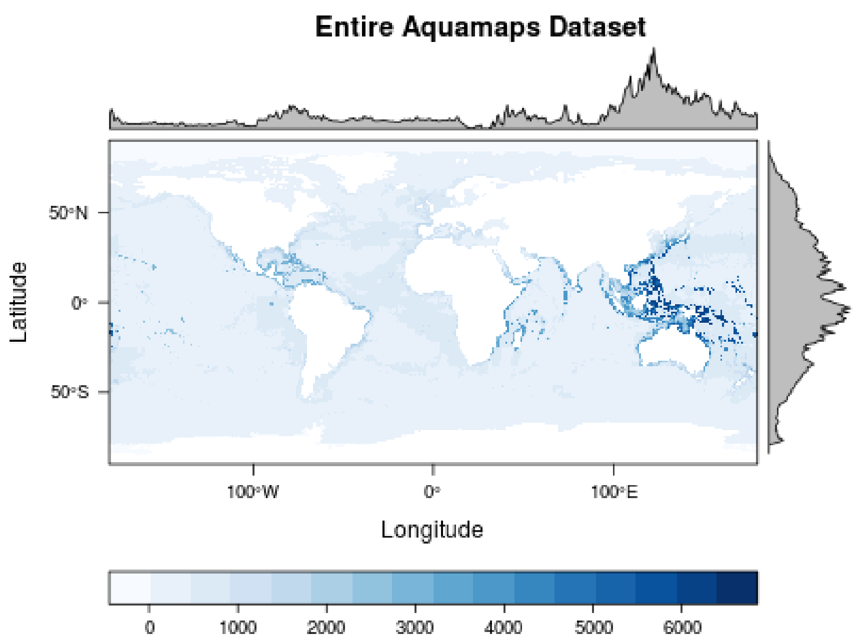
3.0.3 Methods for OHI comparison

3.0.4 Methods for MPA Gap Analysis comparison

4 Figures

4.0.1 Figure 1





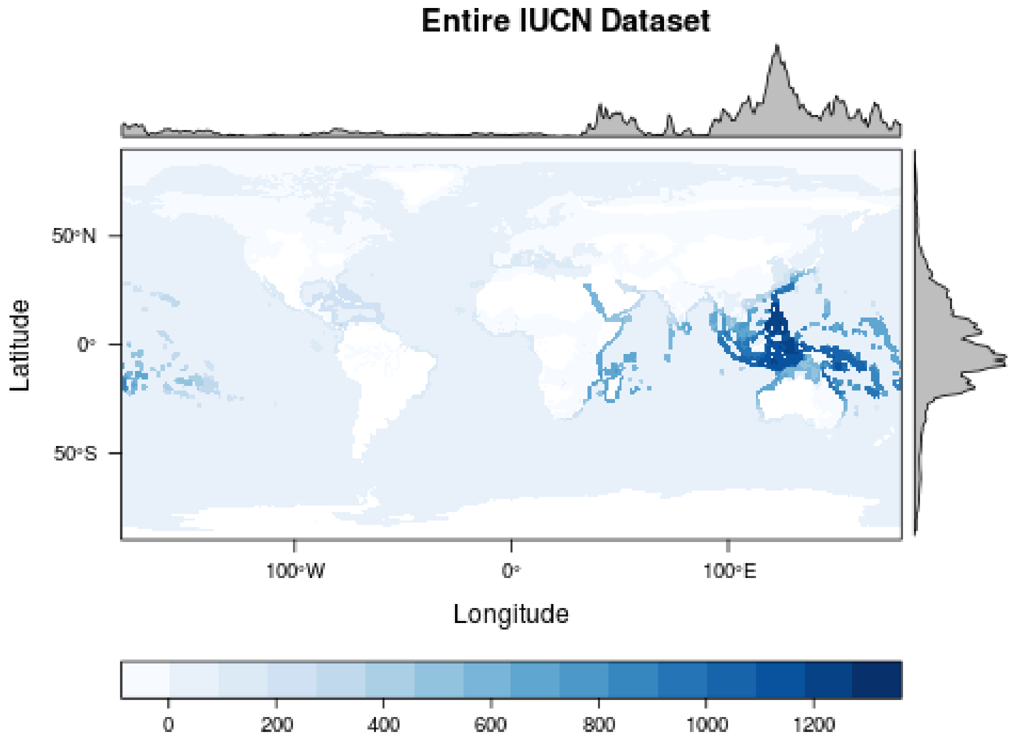


Figure 1(a): Number and proportion of species, listed by taxa, included in each dataset: IUCN, AquaMaps, or both. AquaMaps encompasses a broader range of taxa than IUCN, while IUCN focuses on comprehensively assessing select taxonomic groups, typically at the level of order or family. Overlapping species are dominated by bony fishes (1304 species, primarily tropical taxa) and corals (505 species).

Figure 1 (b, c): Global marine species richness according to (a) AquaMaps dataset and (b) IUCN dataset. The frequency plot to the right of each map shows relative species count per cell at each latitude; 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 frequency plot above each map shows relative species count at each longitude, showing a slight bias in the IUCN dataset away from the Atlantic and eastern Pacific compared to AquaMaps.

4.0.2 Figure 2

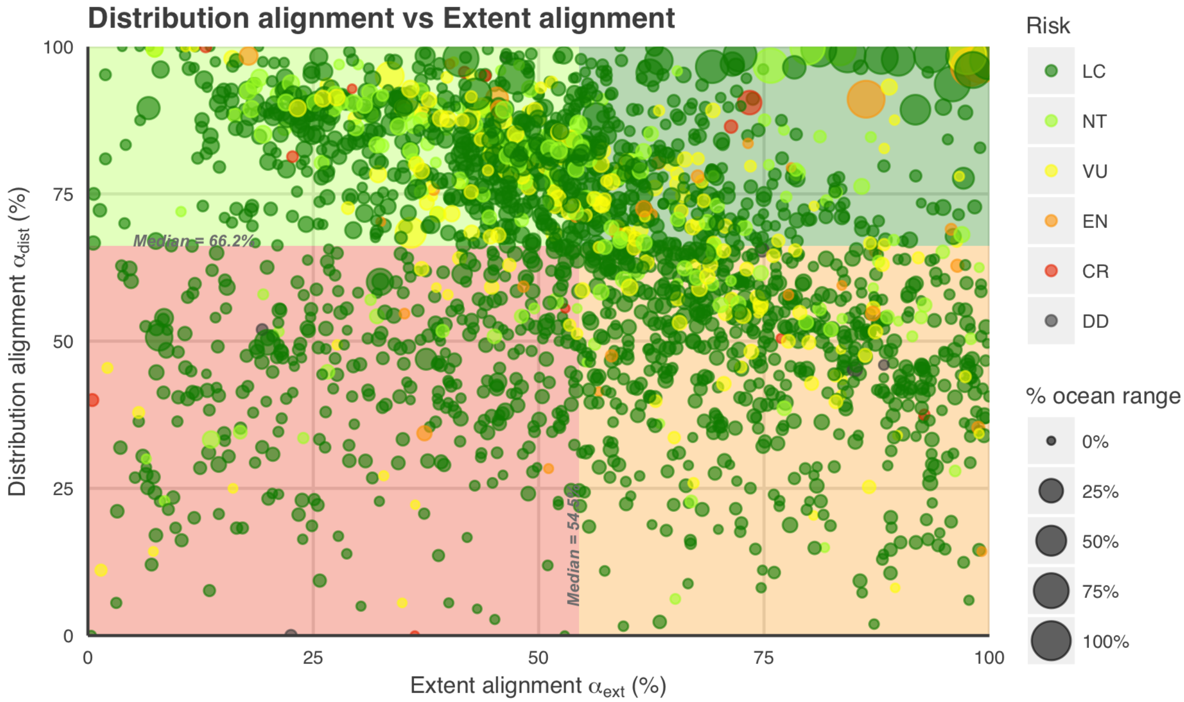


Figure 2 (a): For each paired-map species, we calculated two dimensions of spatial alignment: *distribution alignment*, which we defined as the proportion of the smaller range intersecting the larger range (where on the map); and *extent alignment*, which we defined as the ratio of the smaller range to the larger range (how much of the map). For a species whose distribution is well understood and described in both datasets, we would expect to see a value near 100% for each dimension of alignment.

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 = 400; 18.5 %). The upper left quadrant (quadrant 2) comprises species whose maps agree well in distribution, but disagree in extent (n = 684; 31.6 %). The lower right quadrant (quadrant 3) includes species for which the paired maps generally agree in range extent, but disagree on where those ranges lie - a more worrisome mismatch than that indicated by quadrant 2 (n = 681; 31.4 %). The lower left quadrant (quadrant 4) indicates species for which the map pairs agree poorly in both area and distribution (n = 401; 18.5 %).

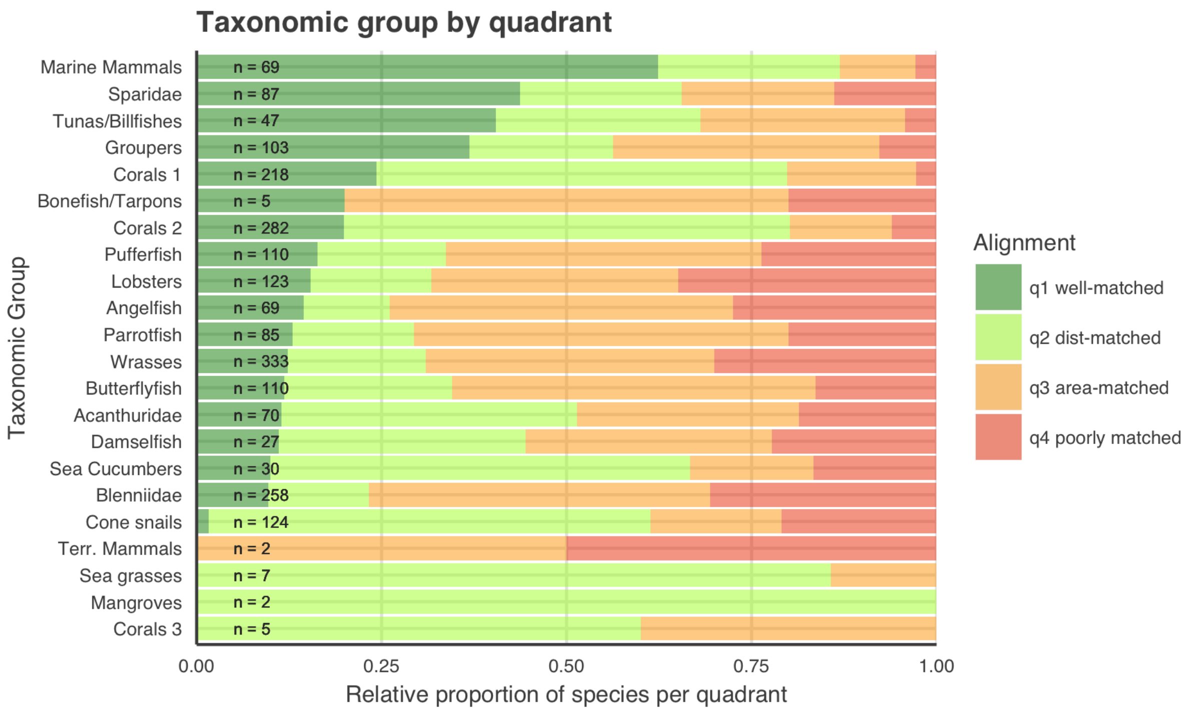


Figure 2 (b): Spatial alignment of paired-map species by taxonomic group. ***REDO with all corals grouped together!*** *Also - is it distracting to use order names for some groups and not others? Some of the groups had pretty long names ("acanthuridae" is way shorter than "surgeonfish, tangs, and unicornfish")*

4.0.3 Figure 3

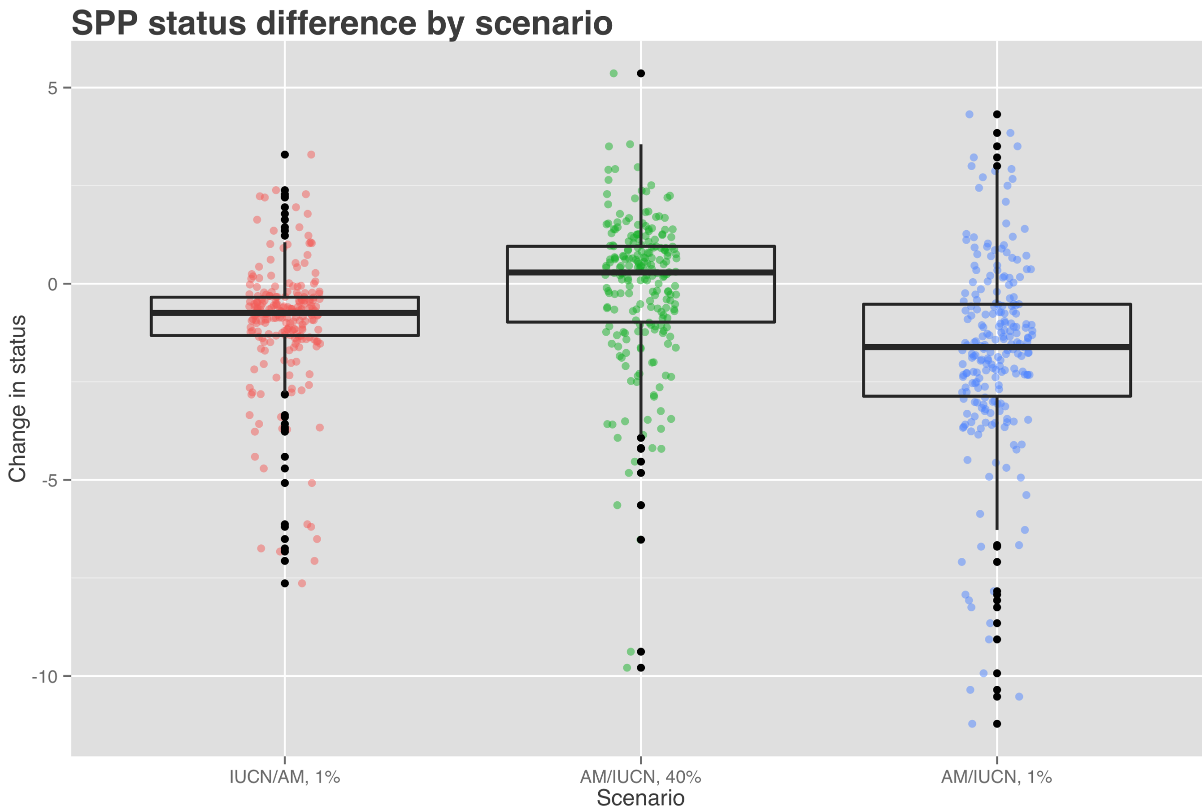


Figure [X] shows the change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios.

Scenario	Priority data source	AquaMaps presence threshold
Scenario 0 (current)	IUCN	>= 40%
Scenario 1	IUCN	> 0%
Scenario 2	AquaMaps	>= 40%
Scenario 3	AquaMaps	> 0%

- Scenario 1 shows the effect of reducing the presence threshold for AquaMaps presence. Reducing the threshold will always increase the apparent range of a species, therefore the slight decrease in average score suggests increased spatial representation of threatened species.
- Scenario 2 shows the effect of preferring AquaMaps data over IUCN, while maintaining the same presence threshold. This will have different effects depending on the species; in general, AquaMaps ranges are smaller than IUCN ranges, so many but not all 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 more likely a combination of both.
- Scenario 3 shows the effect of preferring AquaMaps data over IUCN, while eliminating the presence threshold. Just as a presence threshold of zero in scenario 1 drives a decrease in average score relative to the baseline, the zero threshold in scenario 3 drives a decrease in scores relative to scenario 2. The large decrease seems to indicate that within the set of paired-map species, a zero threshold greatly increases the spatial representation of high-risk species relative to low-risk species.

5 References

Key AquaMaps publications:

Kaschner, K., R. Watson, A. W. Trites, D. Pauly (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*

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

Ready, J., K. Kaschner, A.B. South, P.D Eastwood, T. Rees, J. Rius, E. Agbayani, S. Kullander and R. Froese (2010). Predicting the distributions of marine organisms at the global scale. Ecological Modelling 221(3): 467-478. PDF *Presents AM; assessing AquaMaps against other presence-only species models*

Papers based on AquaMaps:

Jones, M.C., S.R. Dyeb, J.K. Pinnegar and W.W.L. Cheung (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*

Coll, M., C. Piroddi, J. Steenbeek, K. Kaschner, F. Ben Rais Lasram et al. (2010). The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS ONE 5(8): e11842. PDF *used AquaMaps to predict Med biodiversity. Also: Threshold = 0.*

Martin C.S., Fletcher R., Jones M.C., Kaschner K., Sullivan E., Tittensor D.P., Mcowen C., Geffert J.L., van Bochove J.W., Thomas H., Blyth S., Ravillious C., Tolley M., Stanwell-Smith D. (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). PDF *report on marine data sets and data gaps etc, incl both IUCN and AM*

Hurlbert 2007 Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. _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_

Jetz 2008 Ecological Correlates and Conservation Implications of Overestimating Species Geographic Ranges *EOO maps are usually highly interpolated and overestimate small-scale occurrence, which may bias research outcomes*

Pimm 2014 The biodiversity of species and their rates of extinction, distribtuion, and protection. *uses range maps to show biodiversity areas; may use IUCN range maps. Also discusses gaps and possible things that can be done about them.*

Rondinini 2006 Tradeoffs of different types of species occurrence data for use in systematic conservation planning *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.*

García Molinos, Jorge, Benjamin S. Halpern, David S. Schoeman, Christopher J. Brown, Wolfgang Kiessling, Pippa J. Moore, John M. Pandolfi, Elvira S. Poloczanska, Anthony J. Richardson, and Michael T. Burrows. "Climate Velocity and the Future Global Redistribution of Marine Biodiversity." Nature Climate Change advance online publication (August 31, 2015). doi:10.1038/nclimate2769.

Halpern, Benjamin S., Catherine Longo, Darren Hardy, Karen L. McLeod, Jameal F. Samhouri, Steven K. Katona, Kristin Kleisner, et al. "An Index to Assess the Health and Benefits of the Global Ocean." Nature 488, no. 7413 (August 30, 2012): 615–20. doi:10.1038/nature11397.

Klein, Carissa J., Christopher J. Brown, Benjamin S. Halpern, Daniel B. Segan, Jennifer McGowan, Maria Beger, and James E.M. Watson. "Shortfalls in the Global Protected Area Network at Representing Marine Biodiversity." Scientific Reports 5 (December 3, 2015): 17539. doi:10.1038/srep17539.

Selig, Elizabeth R., Catherine Longo, Benjamin S. Halpern, Benjamin D. Best, Darren Hardy, Cristiane T. Elfes, Courtney Scarborough, Kristin M. Kleisner, and Steven K. Katona. "Assessing Global Marine Biodiversity Status within a Coupled Socio-Ecological Perspective." PLoS ONE 8, no. 4 (April 11, 2013): e60284. doi:10.1371/journal.pone.0060284.

- IUCN reference:
  - https://www.conservationtraining.org/course/view.php?id=217&lang=en
  - http://www.iucnredlist.org/technical-documents/red-list-training/iucnspatialresources
  - IUCN Red List - accessed 12/21/2015
- AquaMaps reference, and accessed date?

---

6 Supplemental Information

6.1 Info on data prep

6.1.1 processing AquaMaps

- start with .sql files - three of 'em - how to get 'em?
- turn into .csvs, which columns are critical for this analysis? simplify

6.1.2 downloading and processing IUCN maps

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

6.1.3 creating master species list for co-listed species

- using AquaMaps and IUCN, and IUCN master list, create the big list
- which columns are included?
- adjust this master list to use IUCN SID only for parent - eliminate the whole question of parent/subpop? for the purpose of this analysis

6.2 descriptions of IUCN and AM data sets

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

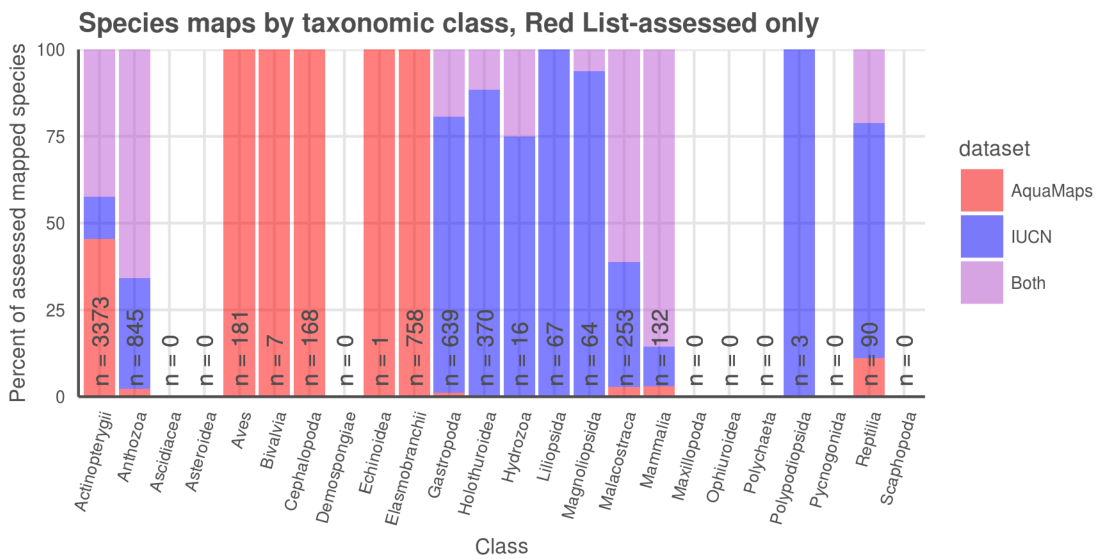
As of December 2015, IUCN had published species distribution maps for 4138 marine species across 24 taxonomic groups. For this analysis, we did not consider IUCN range maps for bird species, as those data are hosted separately by BirdLife International.

As of December 2015, AquaMaps current native distribution maps have been produced for 22889 species.

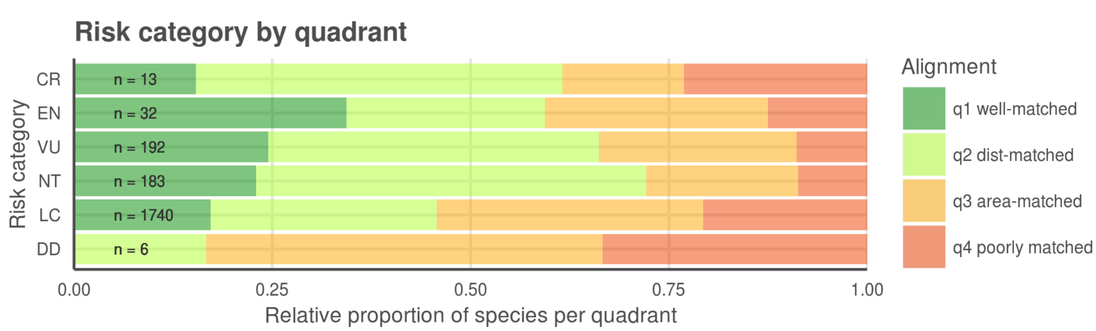
Move to supplementary materials. - BH: For example, as of this writing, IUCN has released no spatial data for class Elasmobranchii (cartilaginous fishes including sharks and rays); and while IUCN offers a large number of maps within class Actinopterygii (ray-finned bony fishes), the available maps include 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.

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.

Red List inclusion:



- *All IUCN-mapped species are also included in Red List species, but only 30% of AquaMaps species.*



- 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*

6.3 illustrative maps for different quadrants and different mechanistic problems

a few codes to indicate likely problems?

- DC = depth clipping; IUCN extents go beyond reasonable depth for the species
- DX = data excluded; AM excludes some observations that appear to be used for IUCN
- DP = data poor (AquaMaps acknowledges data poor status? or use less than 10 data points as criteria)
- NX = needs evaluation by expert; AM model likely predicts areas where species is unlikely to be found based on observed occurrences - does AM model reflect same as AM suitable habitat?

Q2:

species	quad	group	codes	notes
Conus episcopatus	q2	cones	DP, DX	d-match around Coral Triangle, but IUCN shows much more
Ctenochaetus binotatus	q2			IUCN extents beyond AM
Naso vlamingii	q2	acanthuridae		IUCN extents beyond AM
Thalassoma purpureum	q2	wrasses		IUCN extents beyond AM
Porites nigrescens	q2	corals 2	DP	AM limited to CT; IUCN shows far greater dist, and extents
Conus ammiralis	q2	cones	DP	AM
Conus tessulatus	q2	cones		extents
Acropora sarmentosa	q2	corals 1		extents; also AM shows HI but IUCN does not?
Holothuria fuscogilva	q2	sea cucumbers		extents; also AM shows Mex and Central Am west coast, but
Oculina Varicosa	q2	corals 2		lim to Caribbean; extents - this might be a good close-up ma
Acanthocybium solandri	q2	tunas/billfish		mostly tropical by IUCN; AM shows much farther N and S

Q3

species	quad	group	codes	notes
Conus magnificus	q3	cones		IUCN distribution is more limited than AM distribution)
Abudefduf concolor	q3	damselfish		Baja for AM, southern Central Am for IUCN; maybe a good clo
Centropyge aurantonotus	q3	angelfish		AM shows Caribbean; IUCN shows northern South America, als
Chlorurus perspicillatus	q3	parrotfish		IUCN shows only HI; AM shows dots all over S Pacific and nea
Sarda orientalis	q3	tunas/billfish		limited dists by IUCN
Montastraea franksi	q3	corals 2	DP	IUCN shows just in Caribbean; AM shows all up & down east c

Q4

species	quad	group	codes	notes
Acanthurus nigroris	q4			IUCN dist limited to HI, while AM shows all over S Pacific and CT - A
Praealticus tanegasimae	q4	blenny		IUCN shows limited dist around Japan/Marianas; AM shows much bro
Stethojulis marquesensis	q4	wrasse		IUCN shows only small cluster of islands (Marquesas?) while AM show
Canthigaster leoparda	q4	puffer		IUCN limited spots around a couple of islands; AM shows all througho
pentacheles snyderi	q4	lobster	DP	AM shows all over - sim to depth profiles? while IUCN shows only a co

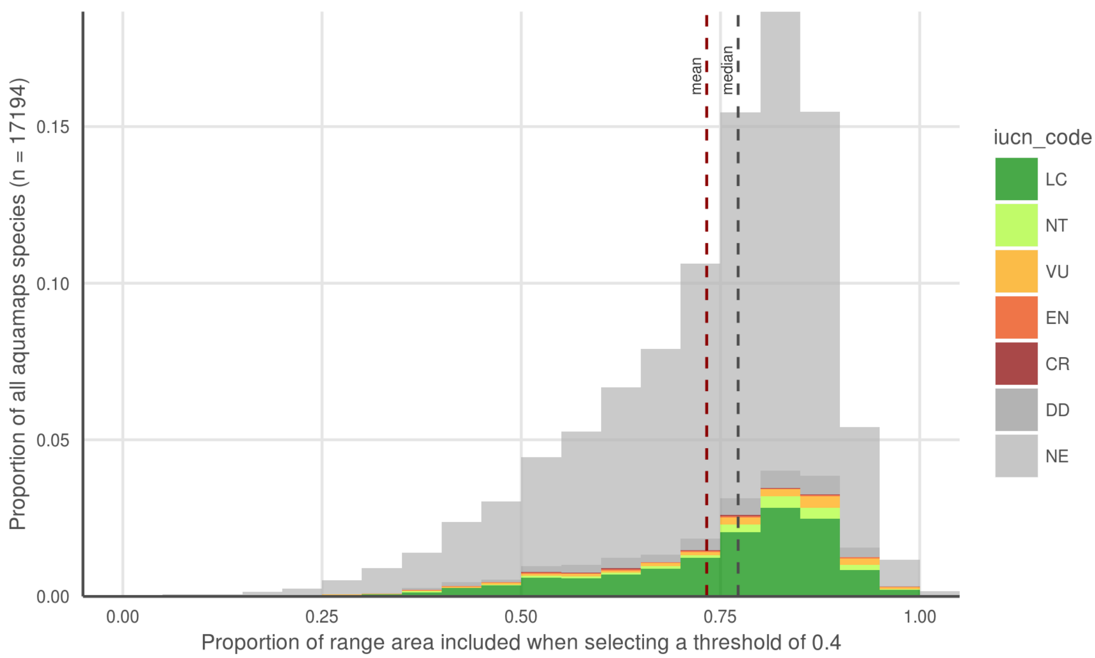
species	quad	group	codes	notes
Nephropsis sulcata	q4	lobster		IUCN shows odd ranges, most of which don't overlap with AM

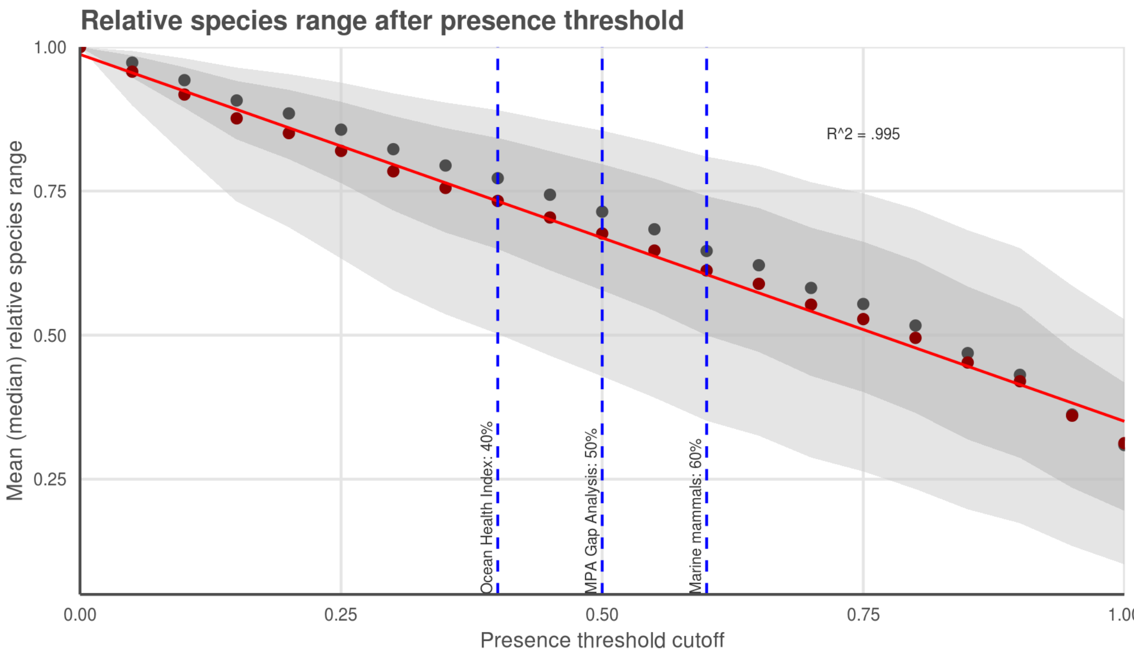
6.4 AquaMaps: Effect of changing "presence" threshold on apparent distribution

6.4.1 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.

6.4.2 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.