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

---

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 and the International Union for Conservation of Nature (IUCN), dominate our understanding of marine species ranges throughout the world's oceans. Together they represent 28,834 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 biodiversity goal and a global analysis of gaps in coverage of marine protected areas - and find significantly different estimates of the status of global biodiversity and effectiveness of conservation depending on which dataset was used. **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 2014 *IUCN, mostly focusing on terrestrial but some marine*), predicting range shifts (Molinos et al. 2015 *AquaMaps*, and setting conservation priorities (Klein et al. 2015 *AquaMaps*).

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; 79.4% of total), such that most global analyses related to biodiversity to date have used AquaMaps (IUCN range map data exist for only 4,027 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 subject 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) - 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. ***BSH suggestion: We used all unique species range data from both sources but substituted one dataset over the other for shared species to highlight the possible consequences of different data use decisions on our understanding of the status of and protection of marine biodiversity.***

---

2 Results and Discussion

2.1 How and why the datasets differ

The IUCN publishes species range maps, bundled by taxonomic groups (REF), based on expert input on spatial boundaries of a given species' "limits of distribution" (REF IUCN 2014) - 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 (AquaMaps 2013)

The methodologies behind 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 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 (Rondinini, 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).

The negative linear relationship indicates 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.

Dividing the map-paired species into quadrants highlights 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. *BSH: 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 in 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

brief intro to implications - where/how long have IUCN/AquaMaps been used for various purposes; then introduce two case studies

- ***From JSL:*** *I think you need a broader discussion of implications. How many times has IUCN/Aquamaps been cited? How long have they been around? What scales are the studies? Then you can muse about how results/interpretations would be different depending on which dataset they used, and reference Fig 3's.*
- *Then, say for example, the OHI uses these datasets together (don't need to get too much into OHI background I think, maybe they don't even need to know about the biodiversity goal?), but say that to try to combine them they use a threshold at 40% since others use this threshold (but check out the SOM for analyses about thresholds)*

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 subset of species in both sources. OHI 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). Most country scores varied by fewer than five points across the three scenarios; a few country scores decreased by as many as 12 points, indicating a more dire assessment of species conservation status.

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. 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.

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 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...

*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? How much of this area will fall within MPAs and how much will fall outside, proportionally? this affects protection ratio. * More area means more species (relatively) will be at least partially included in MPAs so fewer apparent gap species? * Policy implications? Can also tie this back to the predictions of Rondinini et al re: effects of commission/omission errors*

---

4   Conclusions

What can be done to address the differences between these two important marine species distribution datasets? We have identified several potential causes of commission errors among these datasets. Understanding these causes points toward a few improvements that may help us better align the predictions of species distributions, increasing their utility for ecology research and conservation actions.

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" (REF https://www.conservationtraining.org/mod/scorm/player.php?a=1412&currentorg=M5L1_ORG&scoid=6492). 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 can 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 the area restrictions are denoted by FAO Major Fishing Areas (REF find a reference in AquaMaps somewhere); however, allowing area restrictions based on biogeographical criteria, e.g. Marine Ecoregions of the World (REF), would likely have 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.

---

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 spatially covering a marine area and classified as IUCN I-IV *(Klein paper, table 1, subset 4 - seems to be the one they focus on the most)*, converting the World Database of Protected Areas (WDPA, REF) dataset to a raster of 0.5° cells indicating

proportion of protected area within the cell *(consider also: filter on STATUS == 'Designated', and filter out those pesky fishery management areas and manatee speed zones. . .)*. To verify our reconstruction we ran the analysis using the AquaMaps dataset available in April 2015 (REF) at a presence threshold of 50%. We then updated the analysis using the most recent AquaMaps dataset (REF?) at 50% threshold

- Possible analyses:
  - run AquaMaps (new data) with threshold at 0%
    * already done in original paper for old data, but redo it to compare to AM + IUCN scenario below
  - run with IUCN maps only in place of AquaMaps?
  - run with AquaMaps and IUCN both, using 0% threshold; which preferred?
- Which aspects of the analysis should we focus on?:
  - overall vs broken down by taxonomic groups? (see paper, figure 1)
  - spatial distribution of gap species (see paper, figure 2) <!–
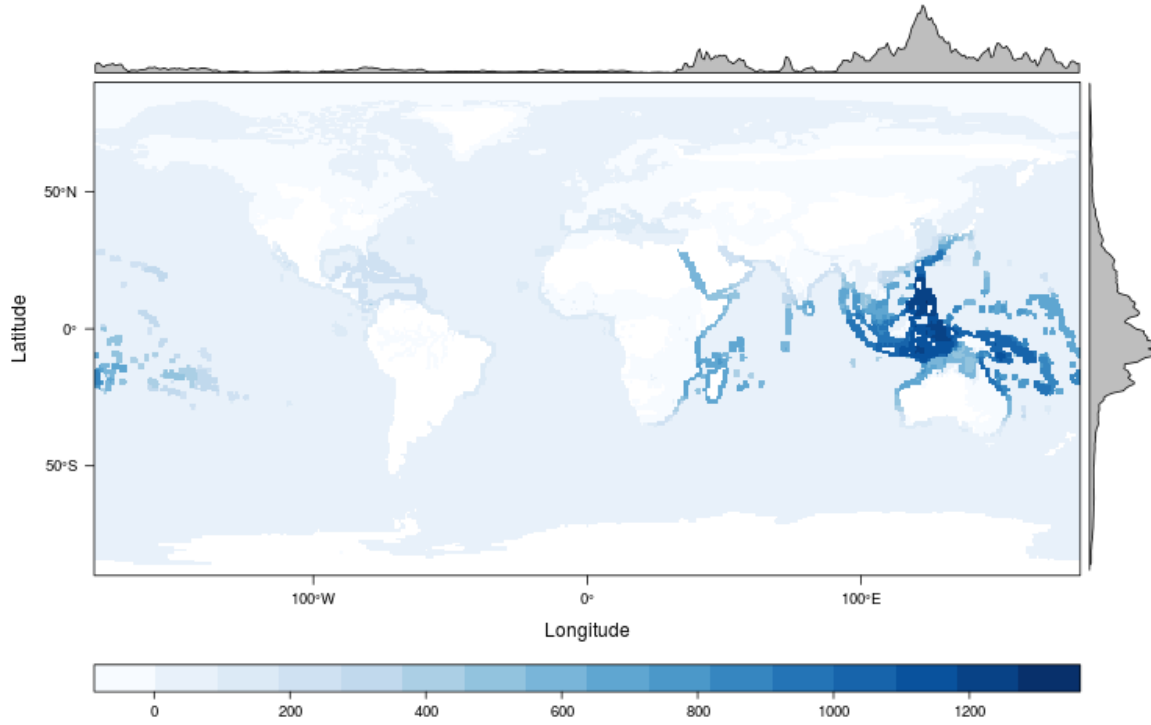
---

6 Figures and captions

still need work on captions; then separate captions from images

---

6.0.1 Fig. 1
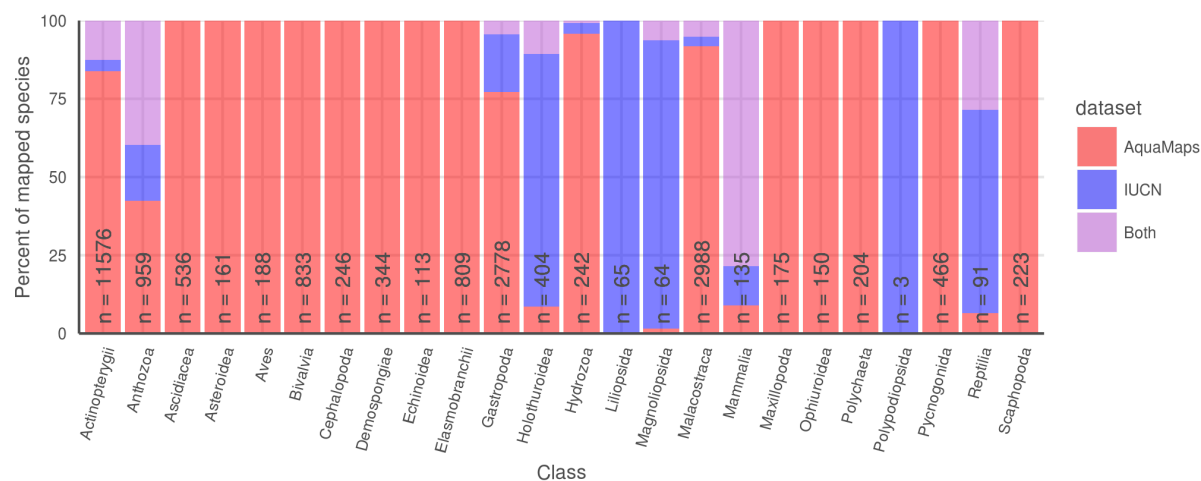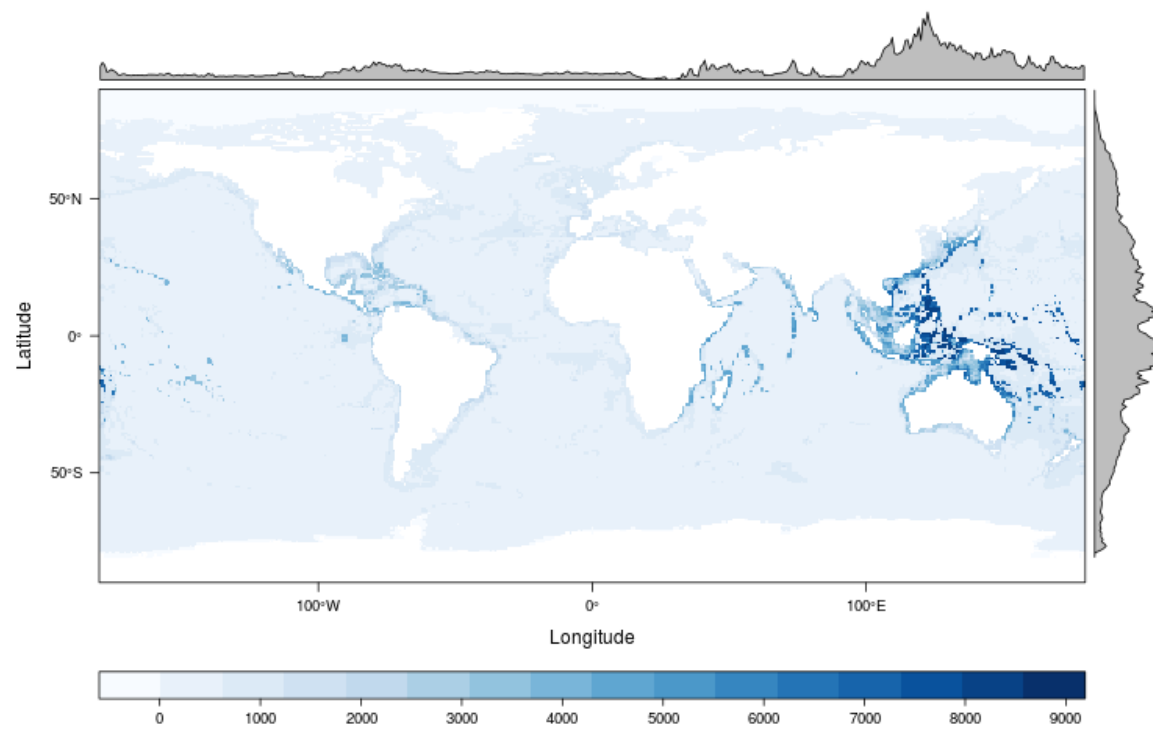
Fig. 1. (A) Number and proportion of species by taxa included in each dataset. AquaMaps encompasses a broader range of taxa than IUCN, while IUCN focuses on comprehensively assessing select taxonomic groups, typically at the level of order or family. Overlapping species are dominated by bony fishes (896 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.
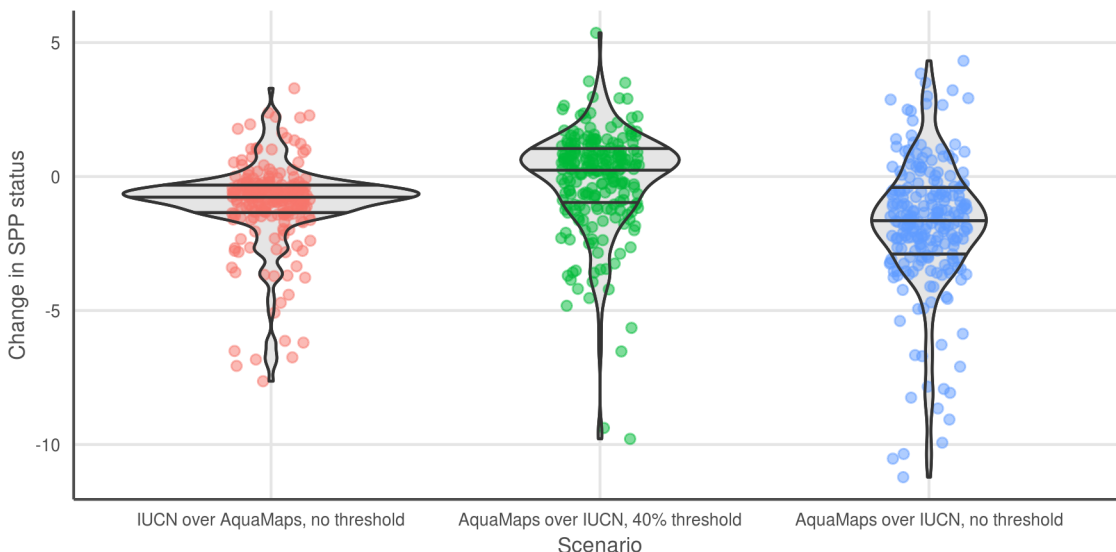
6.0.2 Figure 2

CCO: a lot of this is duplicated in body... tell the story more briefly here; BSH: Yes, make very brief, probably cutting most of this. > **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. *is it distracting to use order names for some groups and not others? Some of the IUCN taxonomic groups had pretty long names ("acanthuridae" is way shorter than "surgeonfish, tangs, and unicornfish") BSH: No, but just be sure to note this difference in the legend.* **NOTE: including the order names might imply the whole order is covered... not necessarily so? Use IUCN taxo groups!**

Fig. 3. 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 probability threshold for AquaMaps presence. Reducing the threshold increases the apparent range of a species; the slight decrease in average score suggests increased spatial representation of threatened species. *CCO: Reference how this more accurately tracks the definition of IUCN "limits of distribution"? BSH: I don't think you need to directly reference this comparison, but these scenarios are helpful to understand what's going on, so they'll need some explanation/implications.*
- Scenario 2 shows the effect of prioritizing 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 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 more likely a combination of both. *CCO: Too much detail here? Just get the main point – probably a mixture of more/less good/bad – can't draw much of a conclusion other than that scores change all over. BSH: You'll need a sentence or two explaining this result, but keep it brief. Extra text, if deemed useful, can be put into the SOM.*
- Scenario 3 shows the effect of prioritizing 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 indicates 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. *similar note as scenario 2*

---

7 References

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

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 *review of marine data sets and data gaps etc, incl both IUCN and AM as well as many others*

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?

---

8 Supplemental Information

8.1 Info on data prep

8.1.1 processing AquaMaps

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

8.1.2 downloading and processing IUCN maps

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

8.1.3 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

8.2   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 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.

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:
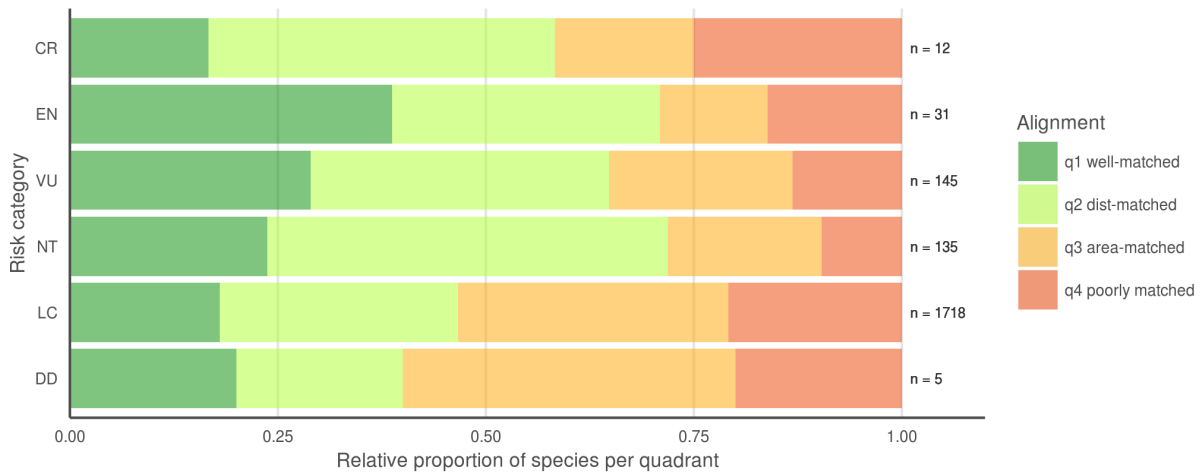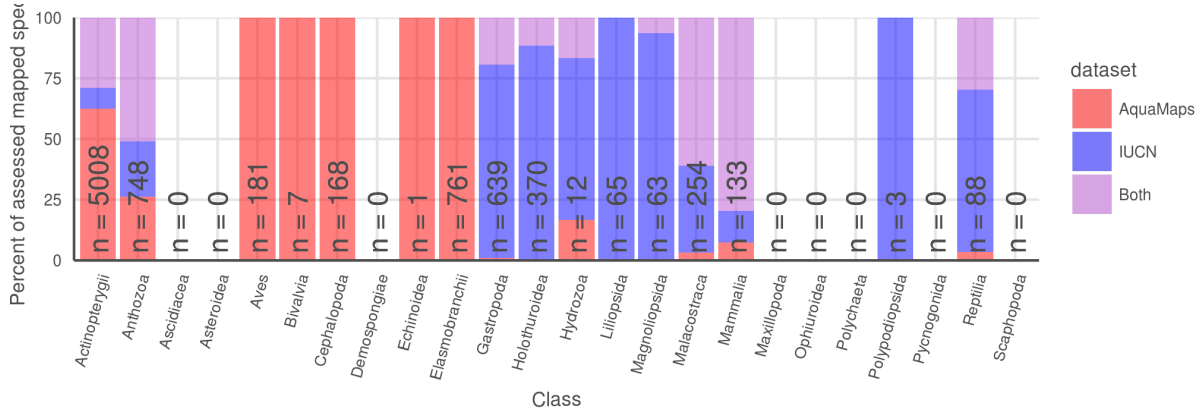
- Breaking down the quadrants by IUCN extinction risk categories (FIG 3c), we found that species

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%)

8.3   illustrative maps for different quadrants and different mechanistic problems
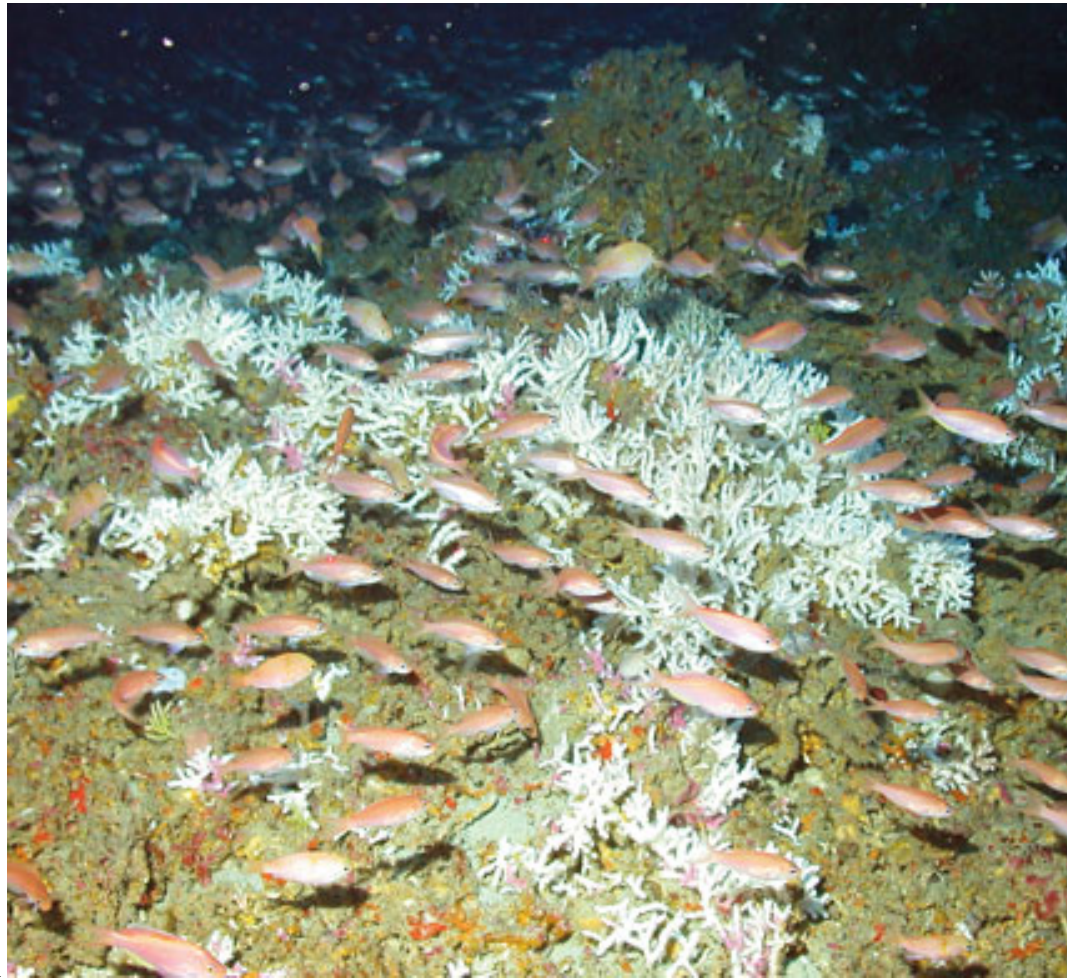
Oculina varicosa - Ivory tree coral

- 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).
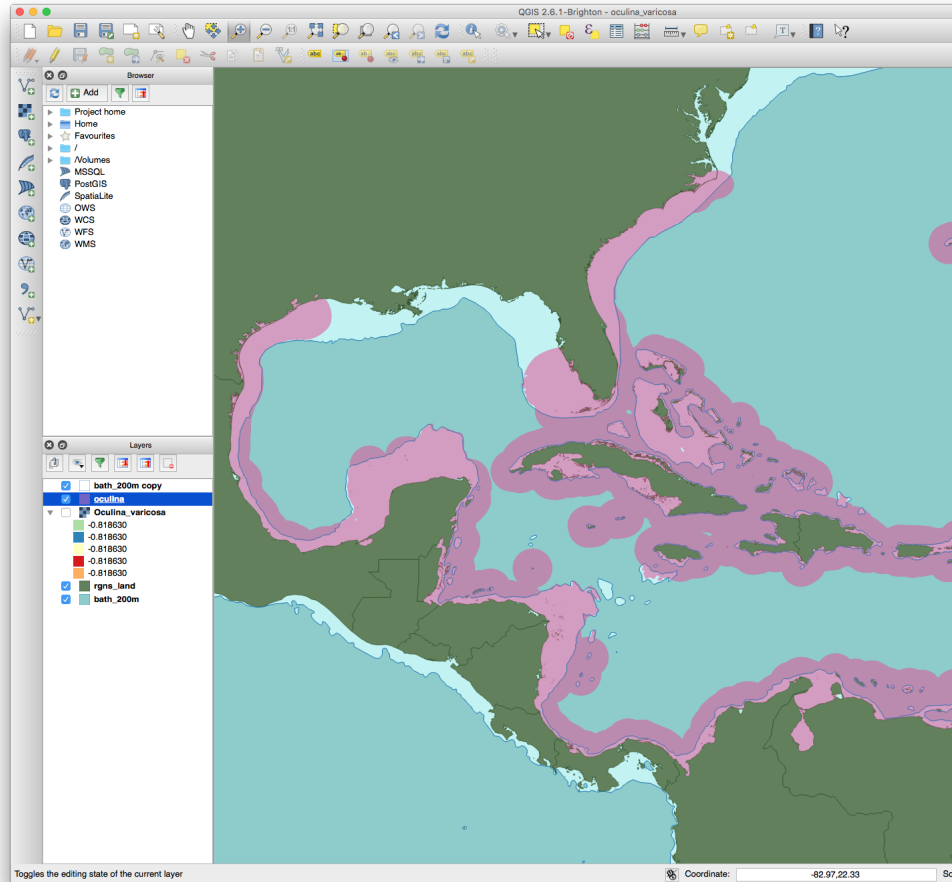
plotted against a 200 m bathymetry line:

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

8.4.1 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.

8.4.2 Figure 5 (a, b):
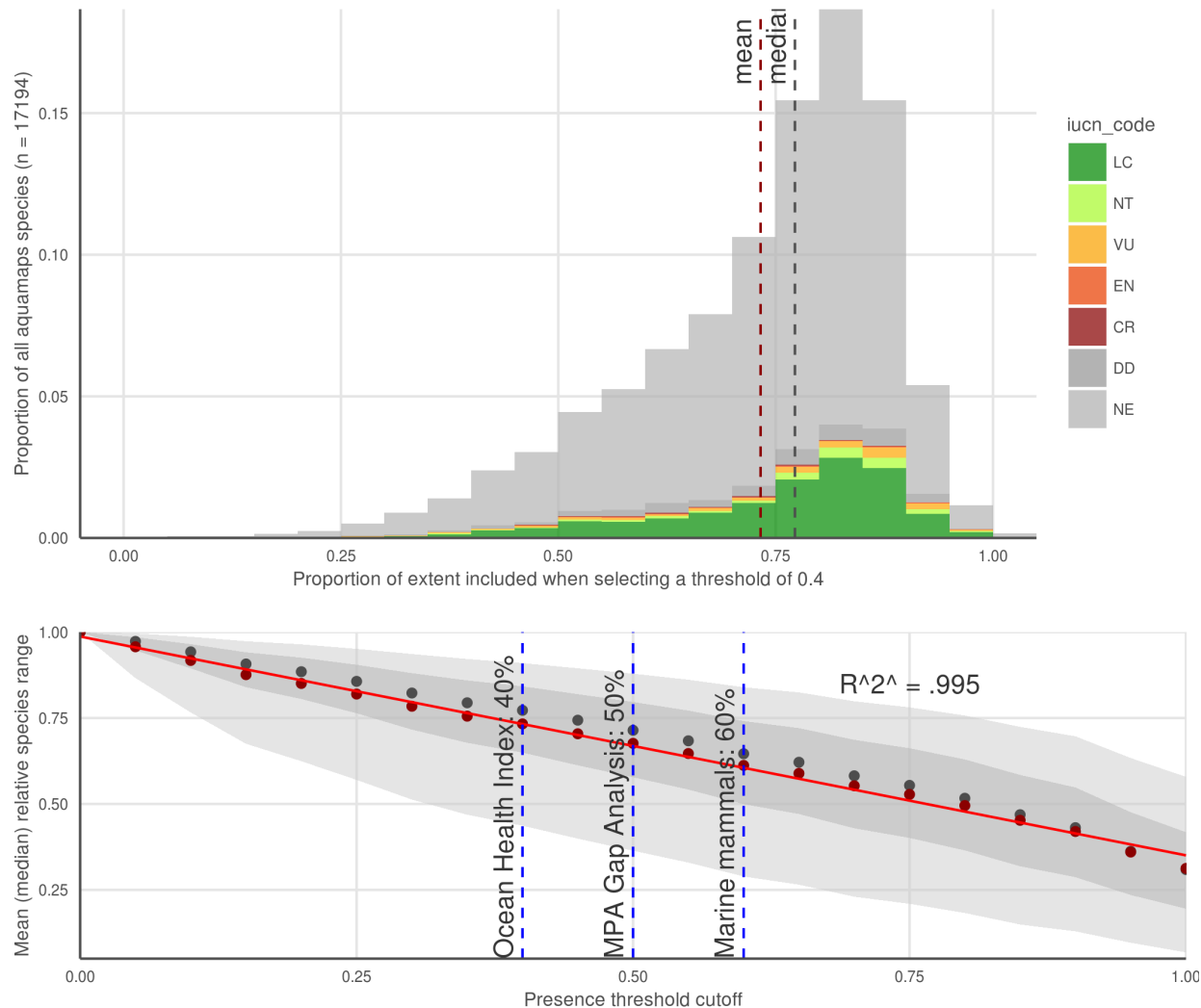
AquaMaps distribution map extent remaining after applying a presence threshold. (a) A 40% threshold applied to all species in the AquaMaps dataset shows a mean loss of XXX, with a wide distribution in which some species lose nearly all of their apparent range. (b) Mean (median) remaining extent at increments 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.

18