Aligning species range data to better serve science and conservation

Casey O'Hara, Jamie Afflerbach, Ben Halpern

0.1 Abstract

get to this later...

0.2 Significance

subset of the abstract

---

1 Introduction

Mapping and predicting species' distributions is fundamental to the sciences of ecology, biogeography, and conservation, among many others. Knowing where individuals of a species are, and what allows them to persist there, provides foundational information for understanding species ranges, how best to protect and manage species, and how they may respond to increasing human impacts and a changing climate. A rich literature exists tackling the many dimensions of these questions.

One very important outcome of this body of science is the various compiled databases of species range maps. In the oceans there are now two such global repositories – Aquamaps (REF) and International Union for Conservation of Nature (IUCN) species distribution maps (REF) – that 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 and trends (REF), predicting range shifts (Molinos et al. 2015), and setting conservation priorities (Klein et al. 2015). The two data collections ostensibly describe the same information, but significant differences in methodology, intent, and taxonomic and geographic sampling could lead to dramatically different understandings of our marine ecosystems and resulting policy and conservation recommendations

Importantly, biases in taxanomic or spatial coverage of a dataset could shift management or conservation actions towards places or species that aren't actually the most in need. [elaborate a bit?]

To understand the implications of similarities and differences between the two datasets, we examined how they compare in their representation of global spatial distribution of species, and for the relatively small number of species represented in both datasets, how closely their spatial distribution maps align. We then reanalyze two recent studies using these datasets based on different versions of the data to illustrate how and why decisions about which species range data to use affect our understanding of the status of marine biodiversity.

2 Overview of the datasets

We may need to shorten this section to fit this paper into the format guidelines for PNAS. If so, some of this can be pushed to the Suppl. Materials. - BH

The IUCN publishes species distribution maps as collections of shapefiles for a range of taxonomic groups, including both terrestrial and marine organisms. Species spatial distributions are determined by experts based upon known occurrences of the species and expert understanding of range and habitat preferences. These maps are typically based on species occurrence records from databases such as the Ocean Biogeographic Information System (OBIS) (REF) and the Global Biodiversity Information Facility (GBIF) (REF); IUCN provides *general guidelines* ***Do we need to describe these? - BH*** to facilitate consistency in how the maps are generated. The IUCN presents these maps as 'limits of distribution'; they are meant to indicate that a species probably occurs within the defined polygon, but are not meant to indicate that the species is evenly distributed within the polygon. While the polygons roughly define regions of presence/absence, additional attributes provide information on extant/extinct ranges, native/introduced ranges, and seasonality.

IUCN publishes spatial distribution map sets only for taxonomic groups that have been "comprehensively assessed," i.e. in which at least 90% of the species within the taxonomic group have been evaluated. 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. As of December 2015, IUCN had published species distribution maps for 4343 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.

AquaMaps (Kaschner et al. 2013) offers species range maps based on modeled relative environmental suitability. For each assessed species, data from occurrence records such as OBIS and GBIF, published species databases such as FishBase, and expert knowledge are used to generate bioclimatic envelopes for sea surface temperature, salinity, depth, productivity, sea ice concentration, and distance to land. The species' suite of bioclimatic envelopes is compared to a map of environmental attributes on a 0.5 degree global grid, creating a global map of probability of occurrence, with cell values ranging from 0.00 to 1.00. These maps can then be reviewed by experts to fine-tune the parameters to better match known occurrences. As of December 2015, AquaMaps current native distribution maps have been produced for 22889 species. Use of these data to predict species presence or absence requires decisions about which probability threshold to use.

Analyses [briefly describe the analyses done]

- *simple taxanomic comparison (describe very briefly)*
- *spatial overlap comparison (describe very briefly)*
- *redo of OHI biodiversity goal (describe briefly what was done previously, and how that was changed for here)*
- *redo global MPA gap analysis (describe briefly what was done previously, and how that was changed for here)*

3 Results and Discussion

3.0.1 Overlap between assessed species

Due to differences in both methodology and intent, significant differences exist between AquaMaps and IUCN species distribution maps (Fig. 1). Only 2455 species were covered by both datasets (0.1% of total Aquamaps species; `p_both_iucn`% of total IUCN species; Fig. 1, 2). **All IUCN-mapped species are also included in Red List species, but only 22.9% of AquaMaps species.**

3.0.2 Taxonomic distribution between datasets

Either move to brief methods section above, and maybe even shorten - BH: 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 further divided the data by whether species had been evaluated for the IUCN Red List of Threatened Species.

Either cut or move to data description section of methods above. - BH: The IUCN only releases spatial data sets when a taxonomic group (on the scale of order or family) has been comprehensively assessed. As such, spatial data for many taxonomic classes are unavailable, and within a class, the assessed sub-groups may not represent the entire class. ***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.

Move to methods section above. - BH: To compare the spatial representation of the two datasets directly, we first rasterized the IUCN species polygons to the same 0.5° cells as the AquaMaps species maps; species presence within a cell was determined by any non-zero overlap of a species polygon with the cell, and species richness per cell was simply the sum of the species present. For the AquaMaps dataset, we determined per-cell species richness by summing all species with non-zero probability of occurrence. We represented relative distribution of species richness for each dataset by plotting average species count against latitude and longitude. To highlight the differences between the two datasets, we also examined the relative difference in per-cell species richness against latitude and longitude.

The balance of species represented in IUCN spatial data skews toward tropical latitudes compared to the distribution represented in AquaMaps distribution maps. This may be driven by the fact that the taxonomic distribution of IUCN distribution maps focuses heavily on coral reef-associated species (see fig. XXX). While both datasets indicate highest species richness in the Coral Triangle and western Indian Ocean, the AquaMaps dataset shows a relatively larger representation of species in the Atlantic, Caribbean, and eastern Pacific than does the IUCN dataset.

Each dataset offers spatial distribution information for large numbers of species. However, the datasets vary in terms of taxonomic coverage and regional coverage. For spatial assessments of biodiversity, the choice of one dataset over the other is likely to create significantly different results. 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 species maps available. For global scale biodiversity studies, however, the selection of one dataset over the other will entail tradeoffs in spatial coverage, taxonomic breadth, and taxonomic depth. ***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*

3.1 Analysis of spatial alignment by species

3.1.1 Defining spatial alignment between the two datasets

For species described in both spatial datasets, we would expect to see spatial correlation both in the global pattern of species distribution and the extent of species range. Large discrepancies in distribution and range area could indicate species that require further study to consolidate expert knowledge for IUCN and improve the inputs to AquaMaps species distribution models.

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 = 2234). Assigning "presence" values using the same methodology above, we found that for 69.6% of paired-map species, the IUCN distribution maps indicated a larger species range than the range indicated by AquaMaps. This aligns with the expectations and early analysis of Kaschner et al (REF).

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.

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

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

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, indicating near-total inclusion of the smaller range within the larger (*distribution alignment*) and very similar extent of range (*area alignment*).

For each of the 2234 species whose range is described by both datasets, we calculated the distribution alignment and the area alignment, and plotted the species map pairs on a scatter plot of distribution alignment vs area alignment (Figure 1).

We divided the plot into quadrants based on the median values for each axis to differentiate between four different qualities of alignment:

- The upper right quadrant (quadrant 1) comprises species whose maps largely agree (better than median value) in both spatial distribution and the area of described ranges. For these species, the expert-drawn range map and the bioclimatic envelope model produce similar predictions of species distribution. (n = 407; 18.2 %)
- The other quadrants each suggest a different story. Species map pairs that fall within the upper left quadrant (quadrant 2: well-aligned in distribution but poorly-matched in area) indicate species in which the range of the smaller map falls generally within the larger map, but the larger map may include more generous buffers or may include areas unrepresented in the smaller map. (n = 708; 31.7 %)
- The lower right quadrant (quadrant 3: well-aligned in area but poorly matched in distribution) includes species for which the paired maps generally agree in range area, but disagree in locations of those ranges. Disagreement in distribution seems to indicate a greater issue than disagreement in area extent, so we categorized map pairs in this quadrant as being lower quality alignment than map pairs in either of the upper quadrants. (n = 706; 31.6 %)
- The lower left quadrant (quadrant 4: poorly matched in both dimensions) indicates species for which the map pairs fail to agree in both area and distribution. (n = 413; 18.5 %)

Breaking down the quadrants by IUCN extinction risk categories, species with higher extinction risk tend to be better aligned between the two datasets. Well-aligned range map pairs (membership in quadrants 1 and 2) may suggest increased expert scrutiny, though this could be explained as higher perceived risk leading to increased attention, or as increased attention revealing deeper risk. Likely both mechanisms are at play in a feedback loop. In any case, poorly-aligned rangemaps, regardless of extinction risk category, seem to indicate species who could benefit from further expert study.

When broken down by taxonomic group, we found that certain taxonomic groups were far more likely than others to be spatially well-aligned; in particular, wide-ranging pelagic organisms such as marine mammals, tunas, and billfishes were more consistently well-aligned (quadrants 1 and 2) than demersal and reef organisms. Coral species are predominantly found in quadrant 2, indicating that while general global distribution trends seem to be consistent between the two datasets, predicted areas are not well-aligned. AquaMaps accounts for seafloor depth as a limiting factor, highly constraining the predicted habitat for photosynthesizing corals, while IUCN maps for coral species often include wide buffers that would extend far beyond appropriate depth for coral habitat, thus likely overestimating the actual species range. A similar mechanism may also drive poor area alignment of other demersal and reef-associated organisms.

Other hypotheses/case studies to highlight other possible mechanistic differences?

3.2 AquaMaps: Effect of changing "presence" threshold on area representation

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 indicated by IUCN maps. Frequently, studies convert this probability to a simple presence value by assigning a threshold value (XXX 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. 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 approximate the presence criteria used by IUCN. 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.

Effect of selecting different thresholds on total "present" area - linearish, no big steps or drops to indicate sudden change in tradeoff. Do regression - shallow slope indicates that sensitivity is less than 1:1.

At a presence threshold of 40%, as used in the Ocean Health Index Species subgoal, the bulk of AquaMaps species suffer a (XXX to XXX %, median = XXX) decrease in represented range, though some species lose nearly their entire range. Examining the entire species catalog as a single set, a unit shift in presence threshold results in a generally smaller proportional shift in represented area, indicating a moderate and consistent sensitivity to threshold with no surprising tradeoffs to inform the identification of a "correct" threshold. This pattern may not hold true for all subsets of AquaMaps, however, whether subsetting by taxa or by georegion.

---

4 Implications

4.1 Application to OHI

The global Ocean Health Index (Halpern et al. 2012), an index made up of 10 goals, utilizes both of these datasets to inform the Biodiversity goal. The biodiversity goal is made up of two parts; Species and Habitats. 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. A probability threshold of 40% is used to determine presence for AquaMaps data. For those species that are are represented within both datasets, IUCN distribution maps were used as the preferred data source. 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).

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.

Since the Ocean Health Index Species subgoal depends on spatial data from both datasets, the impacts of these threshold and preference changes will be somewhat muted. When IUCN data is the preferred data source, only the non-paired AquaMaps species will be affected by threshold changes; and even when AquaMaps is the preferred source, the non-paired IUCN species will dampen the effect of a threshold change.

4.2 Application to MPA Gap Analysis

yeah still gotta do something on this one

---

5 Conclusion

---

6 References

Key AquaMaps publications:

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

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

Papers based on AquaMaps: ***Use these for section 4?***

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

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

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 nd 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

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.

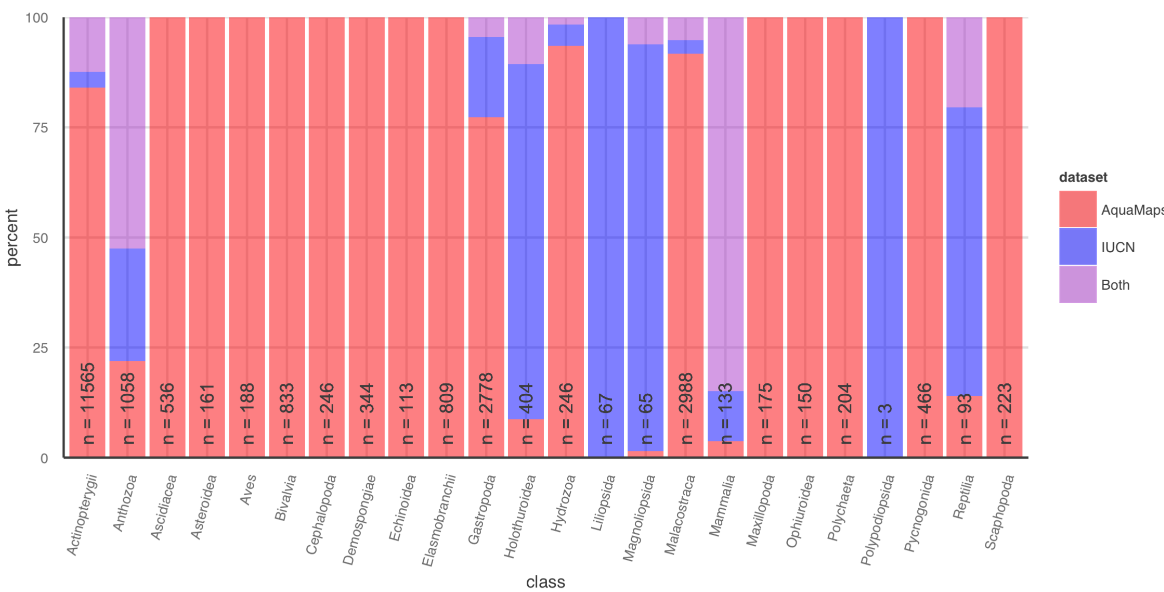
Klein, Carissa J., Christopher J. Brown, Benjamin S. Halpern, Daniel B. Segan, Jennifer McGowan, Maria Beger, and James E.M. Watson. "Shortfalls in the Global Protected Area Network at Representing Marine Biodiversity." Scientific Reports 5 (December 3, 2015): 17539. doi:10.1038/srep17539.

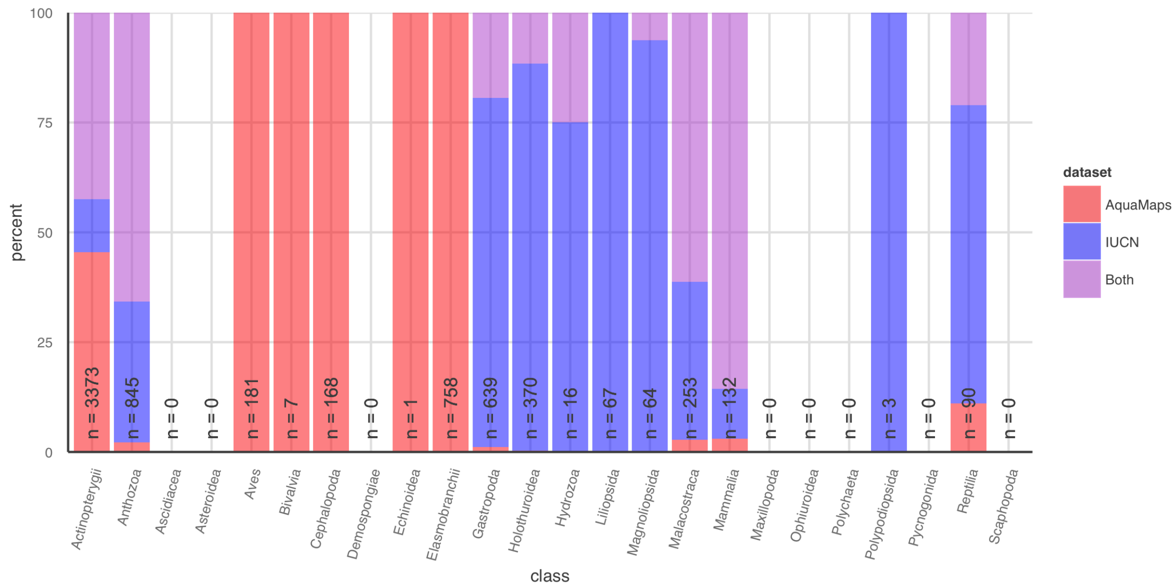
Selig, Elizabeth R., Catherine Longo, Benjamin S. Halpern, Benjamin D. Best, Darren Hardy, Cristiane T. Elfes, Courtney Scarborough, Kristin M. Kleisner, and Steven K. Katona. "Assessing Global Marine Biodiversity Status within a Coupled Socio-Ecological Perspective." PLoS ONE 8, no. 4 (April 11, 2013):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

7 Figures

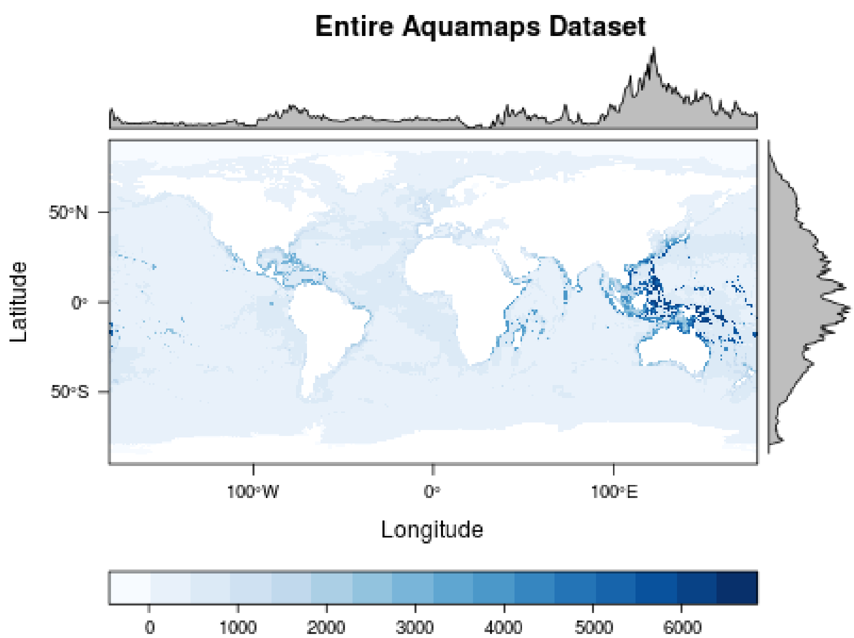
Figure 1 (a, b):





percent
100
75
50
25
0
n = 3373
n = 845
n = 0
n = 0
n = 181
n = 7
n = 168
n = 0
n = 1
n = 758
n = 639
n = 370
n = 16
n = 67
n = 64
n = 253
n = 132
n = 0
n = 0
n = 0
n = 3
n = 0
n = 90
n = 0
Actinopterygii
Anthozoa
Ascidiacea
Asteroidea
Aves
Bivalvia
Cephalopoda
Demospongiae
Echinoidea
Elasmobranchii
Gastropoda
Holothuroidea
Hydrozoa
Liliopsida
Magnoliopsida
Malacostraca
Mammalia
Maxillopoda
Ophiuroidea
Polychaeta
Polypodiopsida
Pycnogonida
Reptilia
Scaphopoda
class
dataset
AquaMaps
IUCN
Both

Figure 2 (a, b):



Entire Aquamaps Dataset
Latitude
50°N
0°
50°S
100°W
0°
100°E
Longitude
0
1000
2000
3000
4000
5000
6000

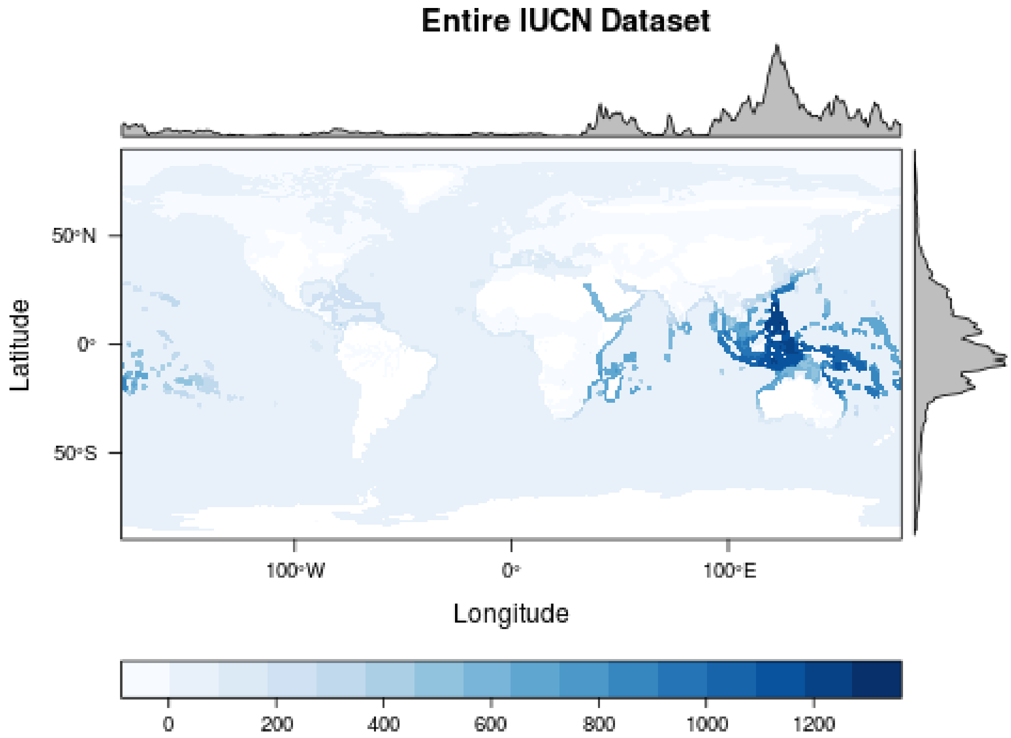

Entire IUCN Dataset
50°N
0°
50°S
Latitude
100°W
0°
100°E
Longitude
0
200
400
600
800
1000
1200

Figure 3:

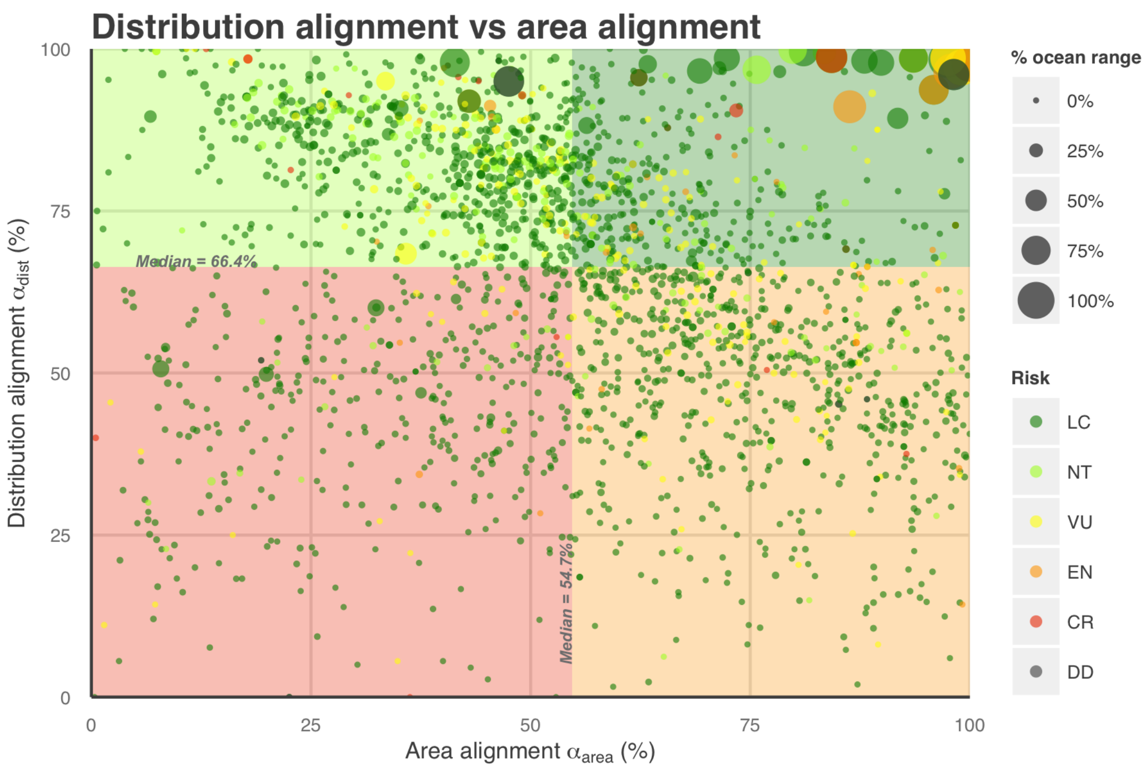
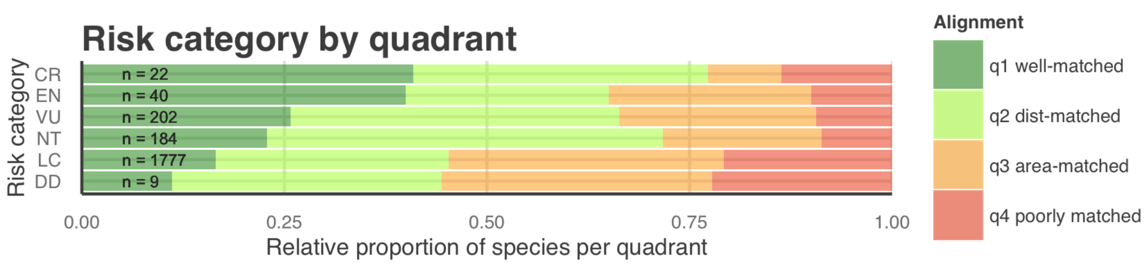


Figure XXX: Distribution alignment vs area alignment for 2234 species mapped in both Aquamaps and IUCN species distribution maps.

Figure 4 (a, b):



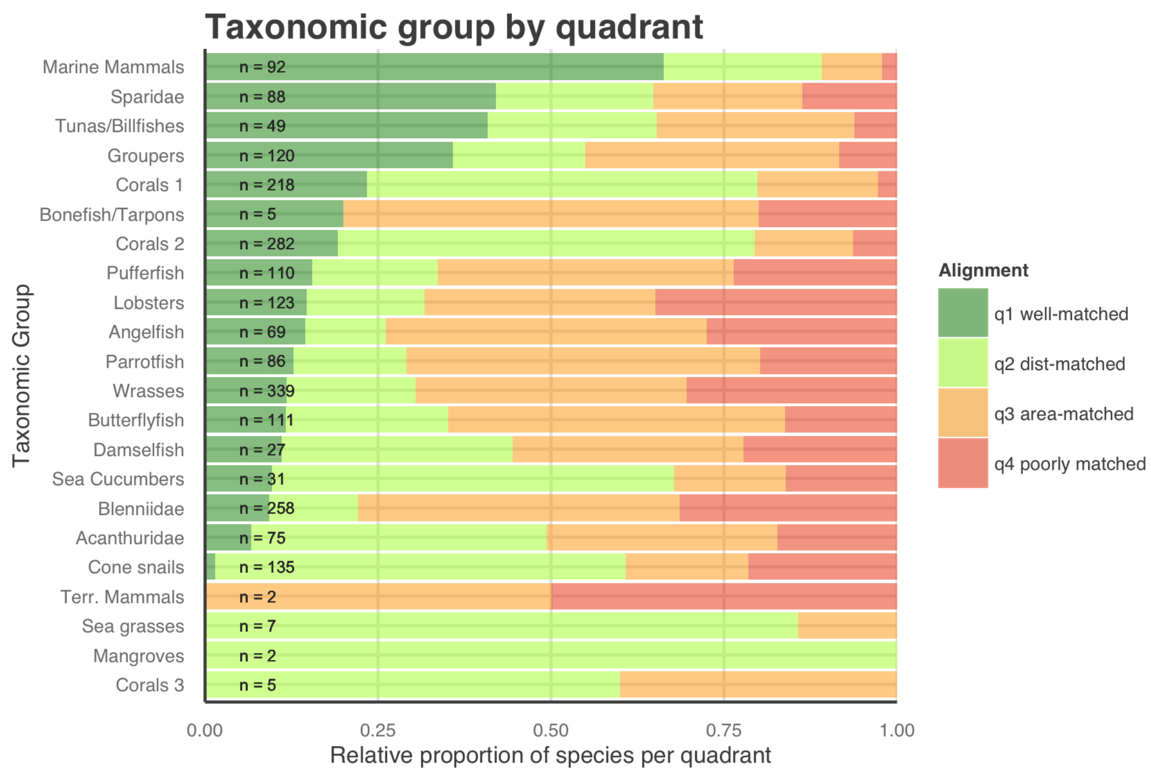
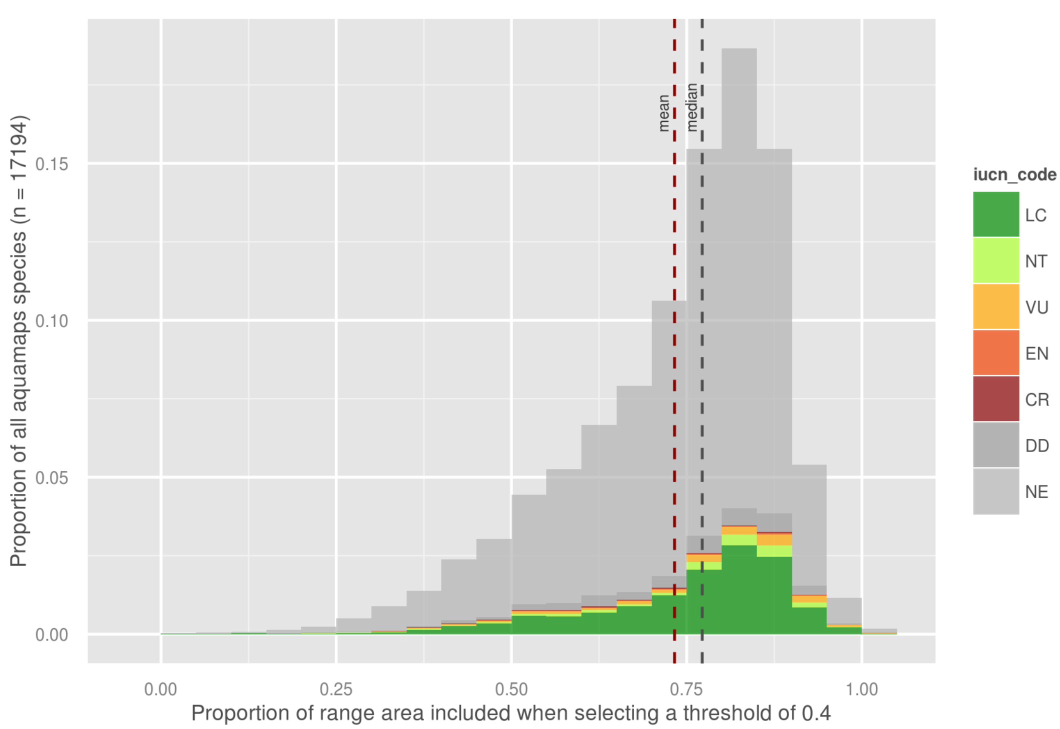

Taxonomic group by quadrant
Marine Mammals
Sparidae
Tunas/Billfishes
Groupers
Corals 1
Bonefish/Tarpons
Corals 2
Pufferfish
Lobsters
Angelfish
Parrotfish
Wrasses
Butterflyfish
Damselfish
Sea Cucumbers
Blenniidae
Acanthuridae
Cone snails
Terr. Mammals
Sea grasses
Mangroves
Corals 3
n = 92
n = 88
n = 49
n = 120
n = 218
n = 5
n = 282
n = 110
n = 123
n = 69
n = 86
n = 339
n = 111
n = 27
n = 31
n = 258
n = 75
n = 135
n = 2
n = 7
n = 2
n = 5
Taxonomic Group
0.00
0.25
0.50
0.75
1.00
Relative proportion of species per quadrant
Alignment
q1 well-matched
q2 dist-matched
q3 area-matched
q4 poorly matched

Figure 5 (a, b):


mean
median
Proportion of all aquamaps species (n = 17194)
0.00
0.05
0.10
0.15
0.00
0.25
0.50
0.75
1.00
Proportion of range area included when selecting a threshold of 0.4
iucn_code
LC
NT
VU
EN
CR
DD
NE

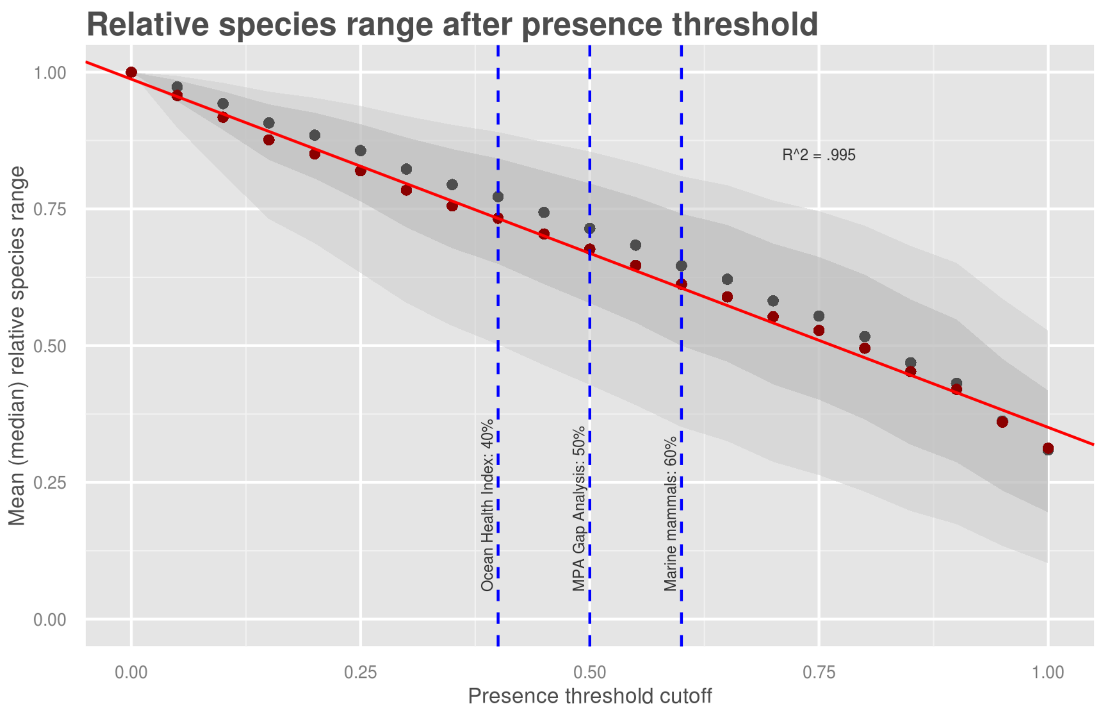


Figure 6

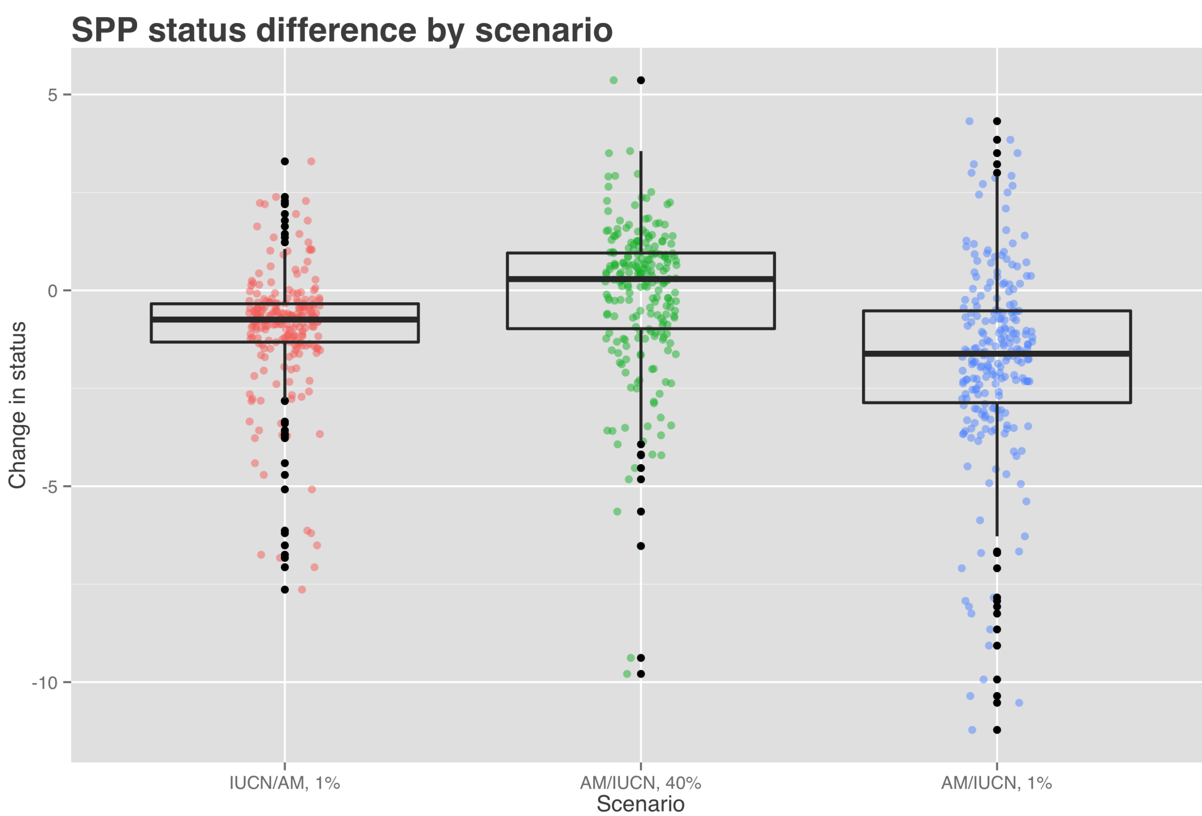


Figure [X] shows the change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios.

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