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 exist, 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 tackles the many dimensions of these questions.

One very important outcome of this body of science is the various compiled databases of species distribution maps. In the oceans there are now two such global repositories *(if we define it narrowly; OBIS is global species distribution, but not quite range maps)* – Aquamaps predicted distribution maps (Kaschner et al. 2013) and International Union for Conservation of Nature (IUCN) range 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 (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 data collections 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 of a dataset could shift management or conservation actions *towards places or species that aren't actually the most in need (shift management or conservation actions away from places or species that are most in need?).*(maybe Jetz 2008?) False indications of presence (commission errors) overestimate the protection afforded by a marine reserve, while false indications of absence (omission errors) reduce the adequacy of a reserve system by underestimating species distributions. (Rondinini et al. 2006)

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 vector shapefiles gathered into taxonomic groups *that have been comprehensively assessed (> 90% of species evaluated)*. Species extents of occurrence are outlined by experts based upon known occurrences of the species, typically from point locality databases such as the Ocean Biogeographic Information System (OBIS) (REF) and the Global Biodiversity Information Facility (GBIF) (REF), and refined by expert understanding of range and habitat preferences.

IUCN publishes range 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.

AquaMaps develops species distribution maps based on modeled relative environmental suitability. For each modeled species, environmental preferences (e.g. temperature, depth, salinity) are deduced from occurrence records such as OBIS and GBIF, published species databases such as FishBase, and expert knowledge. These environmental preferences are compared to a map of environmental attributes on a 0.5 degree global grid, creating a global map of probability of occurrence.

- *How does AquaMaps select species? based on data available from OBIS/GBIF and Fishbase?*

---

3 Results and Discussion

3.0.1 Overlap between assessed species

Differences in methodology and intent drive significant differences between AquaMaps and IUCN species distribution maps (Fig. 1). Only 2455 species were included in both datasets (0.1% of total Aquamaps species; 0.6% 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

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 extent could indicate species that require further study to consolidate expert knowledge for IUCN range maps and improve the accuracy of 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 = 2166). We found that for 69.8% of paired-map species, the IUCN distribution maps indicated a larger species range than the range indicated by AquaMaps. This aligns with the expectations of Rondinini et al. (2006) that range maps are more likely to falsely predict presence while predicted distribution models are more likely to falsely predict absence.

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 *extent alignment*, which we defined as the ratio of the smaller range to the larger range.

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 similar extent of occurrence (*extent alignment*). Dividing the map-paired species into quadrants based on median values for each axis, we can examine the implications of four different qualities of alignment.

Breaking down the quadrants by IUCN extinction risk categories, 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 deeper risk? Likely both mechanisms are at play in a feedback loop.

Examining quadrants by taxonomic group, we found that certain taxa 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 extents 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 (see SI). A similar mechanism may also drive poor extent alignment of other demersal and reef-associated organisms.

Poorly-aligned rangemaps, regardless of taxon or extinction risk category, indicate species that could benefit from further expert study.

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

Confusion in species names also seems to correlate with spatial misalignment. AquaMaps identifies each species map by binomial as well as a unique species ID code. While the IUCN assigns a unique numeric code to every Red List-assessed species, this code is not used consistently to identify range map polygons. ***NOTE: NEW DATA as of Nov 24? Need to re-extract all the IUCN species :(***

range maps less likely to commit omission errors and more likely to commit commission errors; predicted distribution moderate chance of each compared to range maps and point locality data (based on Rondinini et al. 2006) this in discussion after the distribution/extent analysis

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

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. *For the comparisons above, thresh = 0*

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.

---

4 Implications

4.1 Application to OHI

The global Ocean Health Index (OHI) (Halpern et al. 2012), an 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 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. When IUCN data is 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. But

---

4.2 Application to MPA Gap Analysis

do the analysis...

Predictions:

IUCN overestimates 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? Less pressure to carefully consider MPA design; less resolution on

---

5 Methods

taxonomic distribution comparison

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 then filtered the species list to those that have been evaluated for the IUCN Red List of Threatened Species.

Either cut or move to data description section of methods above. - BH: The IUCN releases spatial data sets when a taxonomic group (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.

global distribution spatial comparison

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.

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.

map pairs comparison

from results/discussion: 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.

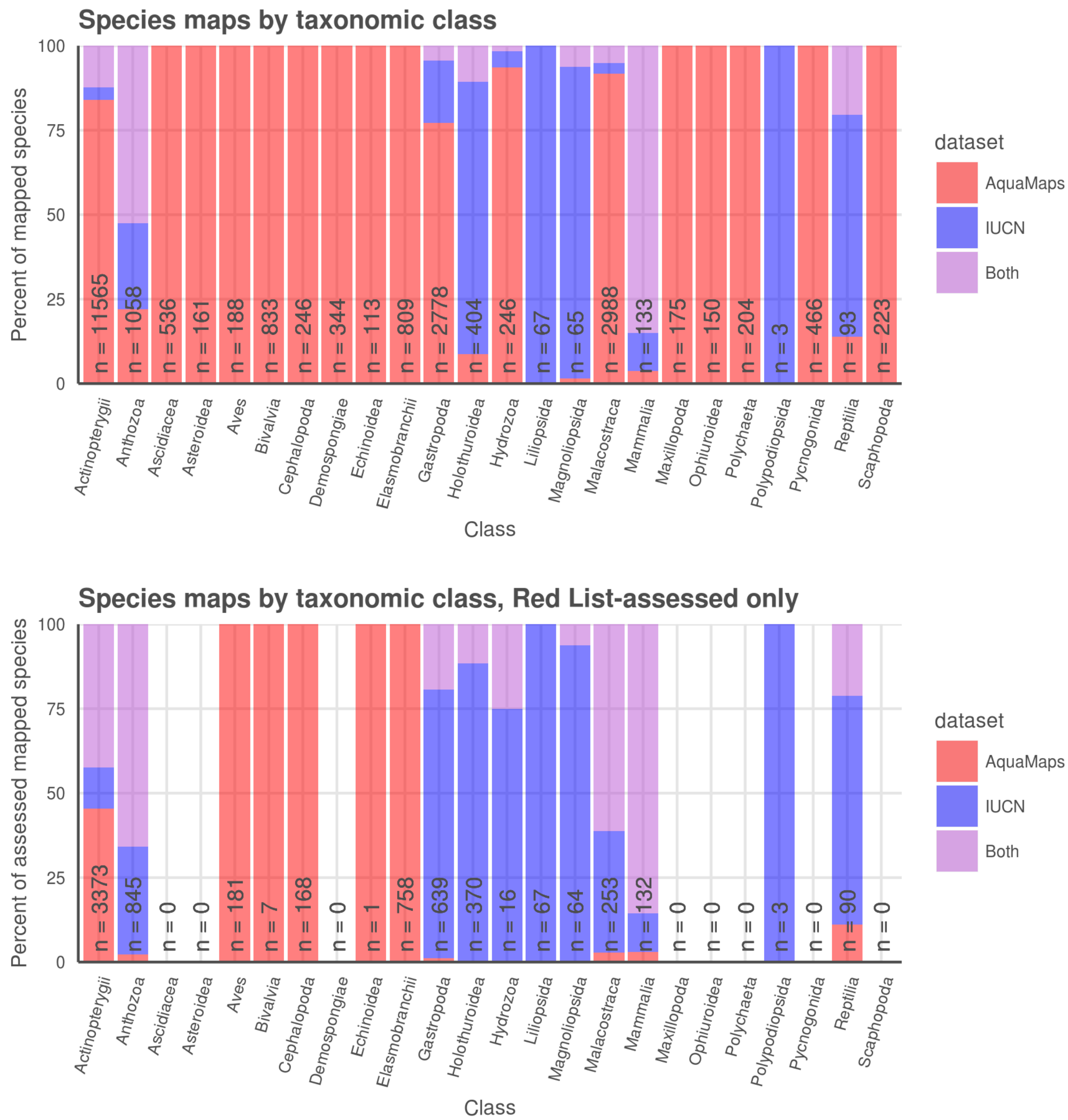
$$\alpha_{dist} = \frac{A_{small \cap large}}{A_{large}} * 100\%$$

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

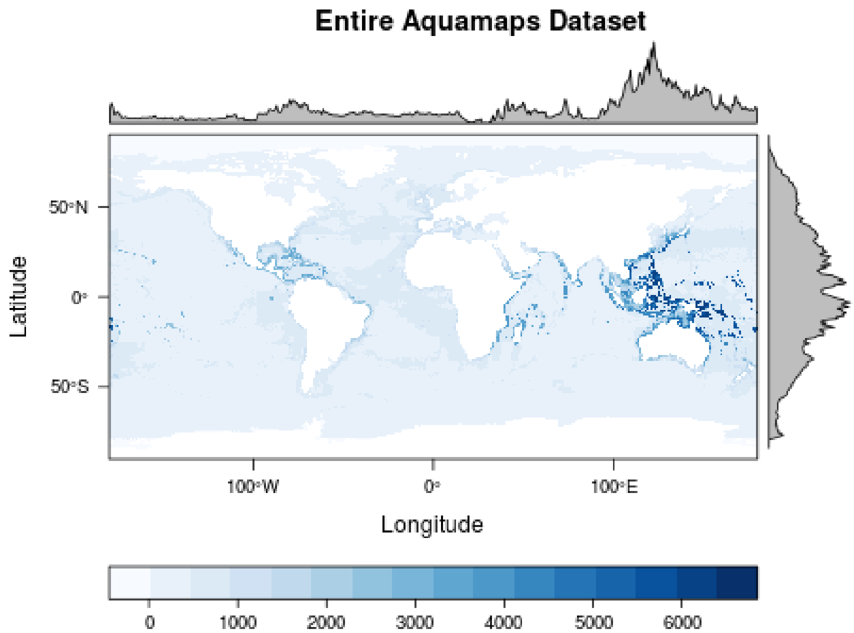
---

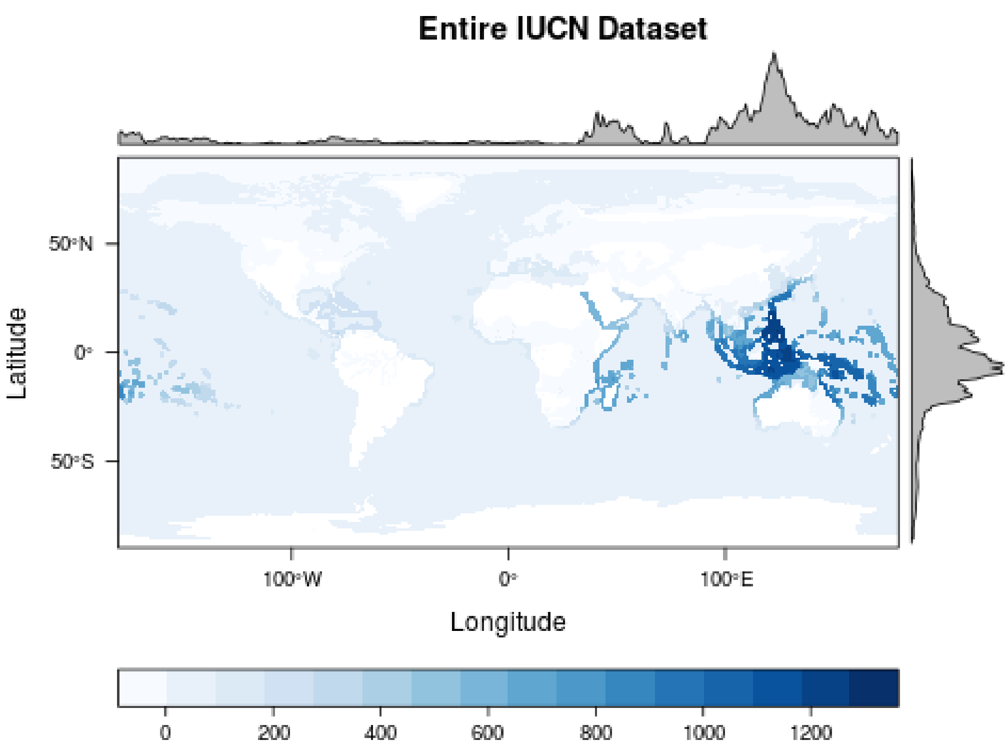
6 Figures

6.0.1 Figure 1 (a, b):



6.0.2 Figure 2 (a, b):





6.0.3 Figure 3:

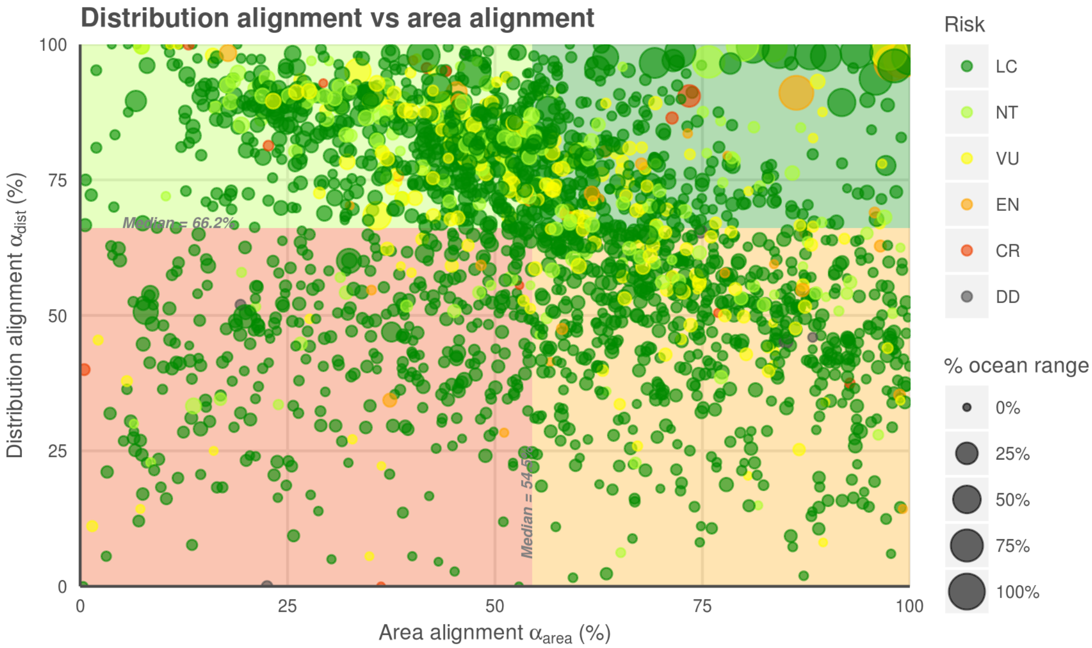
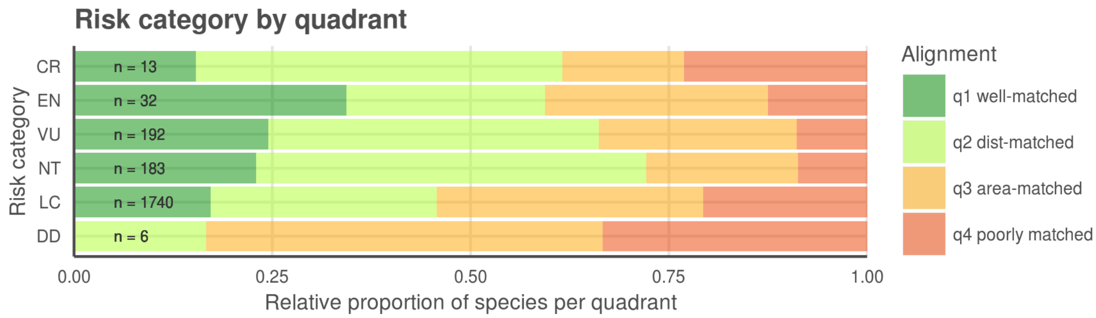
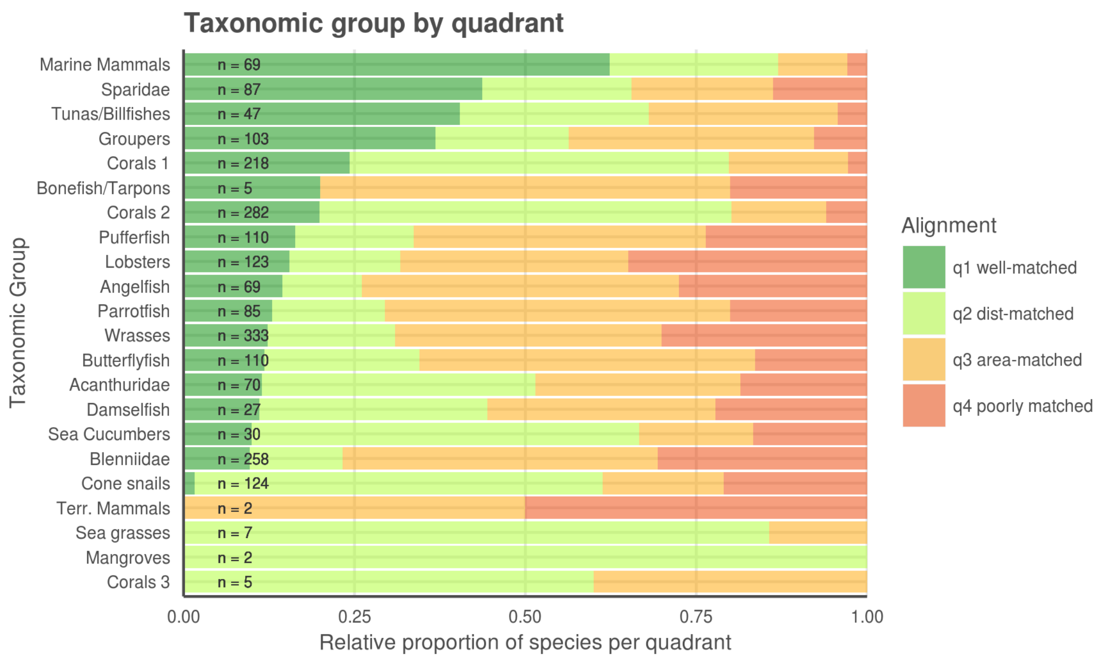


Figure XXX: Distribution alignment vs area alignment for 2166 species mapped in both Aquamaps and IUCN species distribution maps.

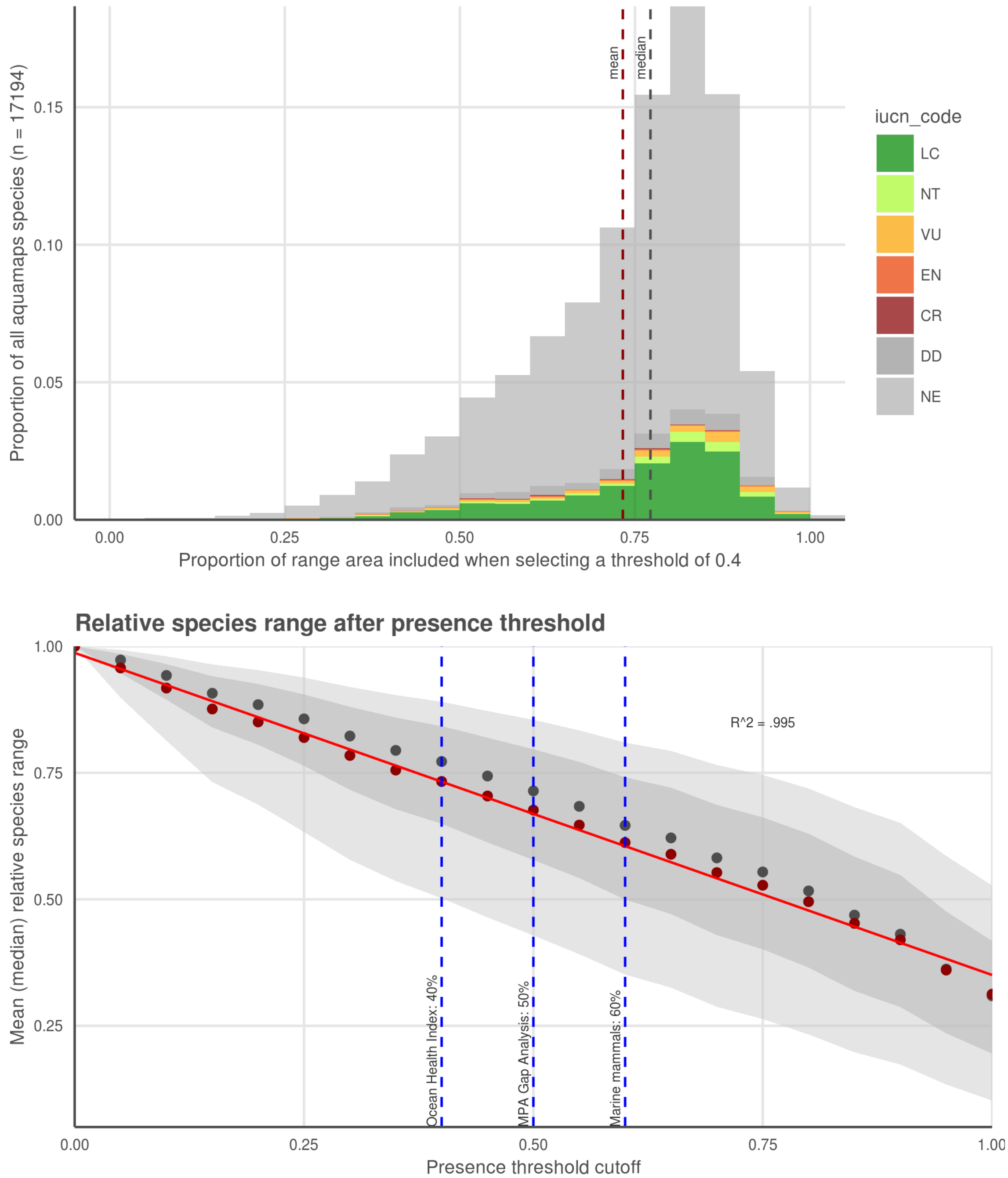
- The upper right quadrant (quadrant 1) comprises species whose maps largely agree (better than median value) in both spatial distribution and the area of described ranges. For these species, the expert-drawn range map and the bioclimatic envelope model produce similar predictions of species distribution. (n = 400; 18.5 %)
- 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 = 684; 31.6 %)
- 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 = 681; 31.4 %)
- 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 = 401; 18.5 %)

6.0.4 Figure 4 (a, b):





6.0.5 Figure 5 (a, b):



6.0.6 Figure 6

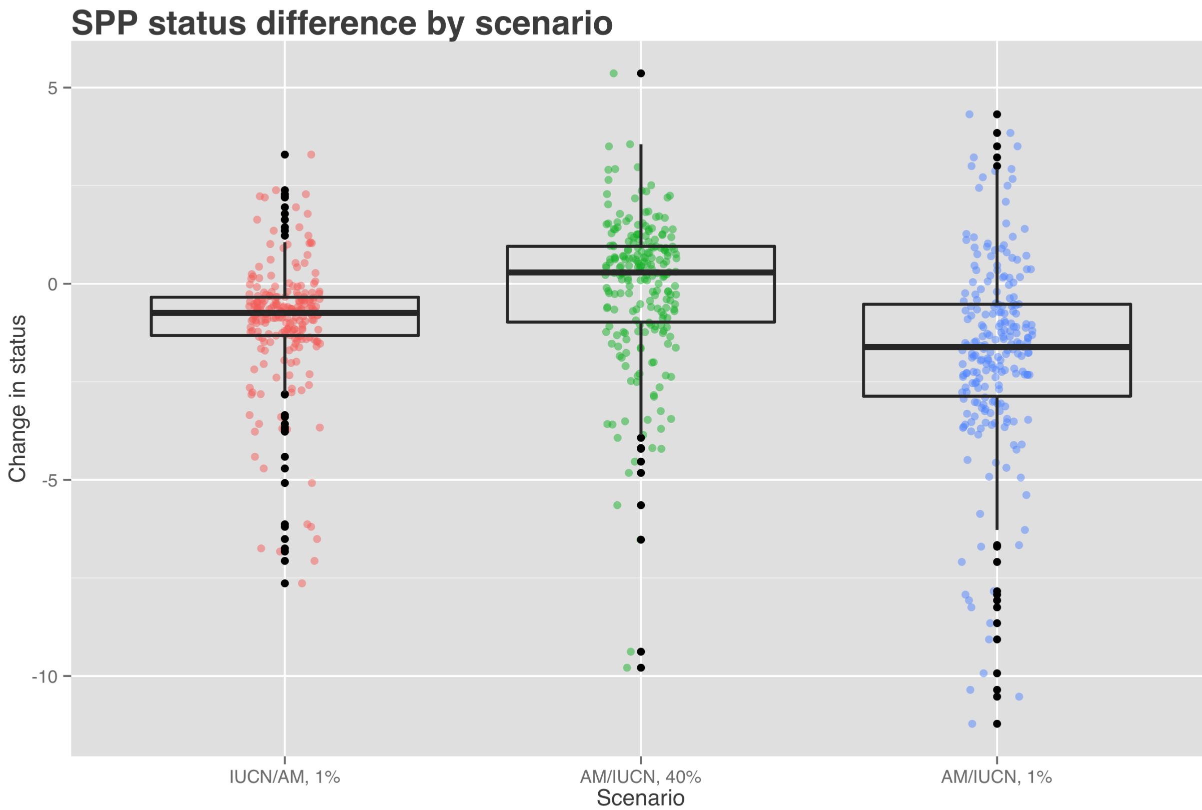


Figure [X] shows the change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios.

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

- 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.

---

7 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?