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

Peer? Dive? Cast a scientific eye? *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. Which species would you expect to find, and which species would be woefully out of place?* 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 where they thrive, 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 important 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 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 or conservation actions away from places or species that are most in need.(maybe Jetz 2008?) Errors of commission (false indications of presence) or omission (false indications of absence) could lead to inefficient or 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. We then subjected two recent marine biodiversity studies - the Ocean Health Index biodiversity goal (Halpern et al. 2012) and a global MPA gap analysis (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.

2 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.

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 while AquaMaps modeled distribution maps are more likely to underpredict species presence.(REF and REF)

3 Results and Discussion

3.0.1 Taxonomic distribution between datasets

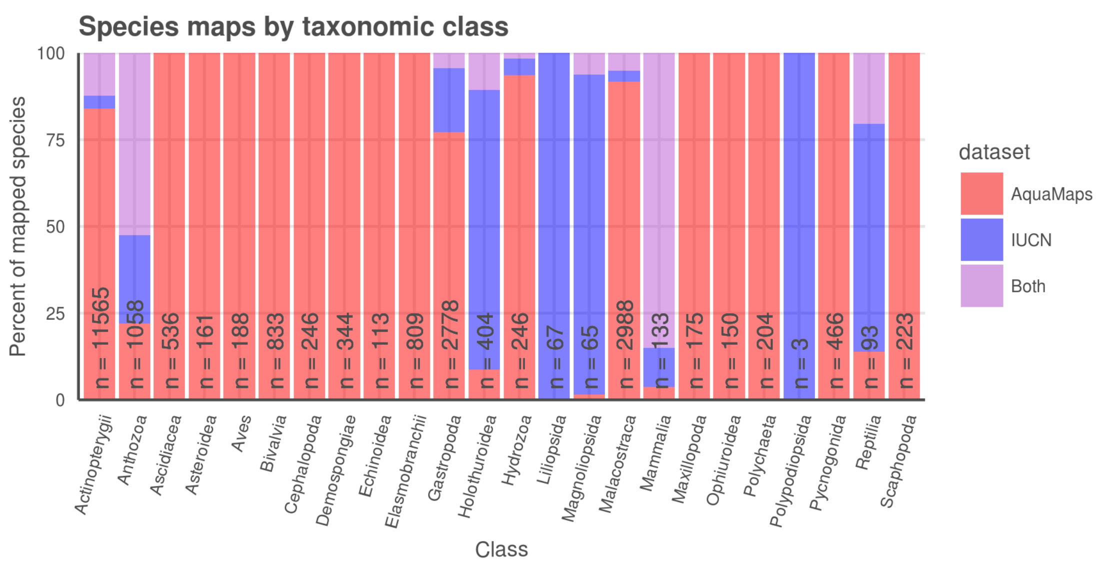
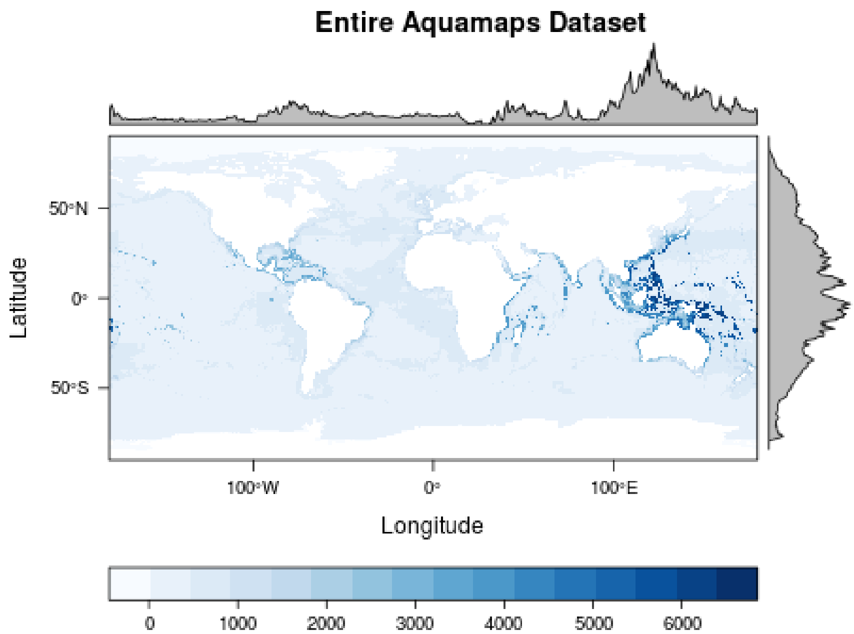


Figure 1(a): Number and proportion of species, listed by taxa, included in each dataset: IUCN, AquaMaps, or both.



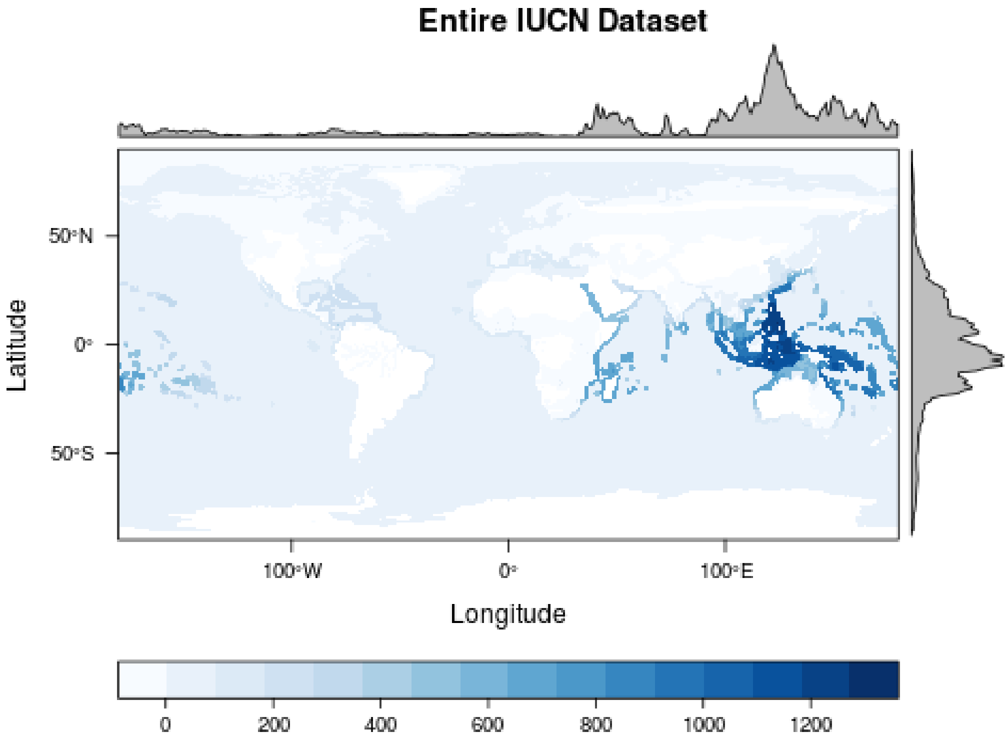


Figure 1 (b, c): Global marine species richness according to (a) AquaMaps dataset and (b) IUCN dataset.

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 2297 species included in both datasets (10% of AquaMaps species; 55.5% of IUCN marine species; Fig. 1, 2).

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.(FIG) This skew likely reflects the fact that the IUCN dataset focuses more heavily on coral reef-associated taxa than does the AquaMaps dataset (see FIG).

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 mapped species 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.*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)

3.0.2 Defining spatial alignment between the two datasets

The IUCN and AquaMaps spatial datasets share 2166 species in common. If we were to examine 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).

In 69.8% of cases, the IUCN range map indicated a larger species range than the AquaMaps map. This finding concurs with the general expectation that geographic range maps (e.g. IUCN) are more likely to over-predict presence than predicted distribution models (e.g. AquaMaps), while predicted distribution models are more likely to over-predict absence.(Rondinini et al., 2006) *errors of commission vs errors of omission, essentially type I and type II errors*

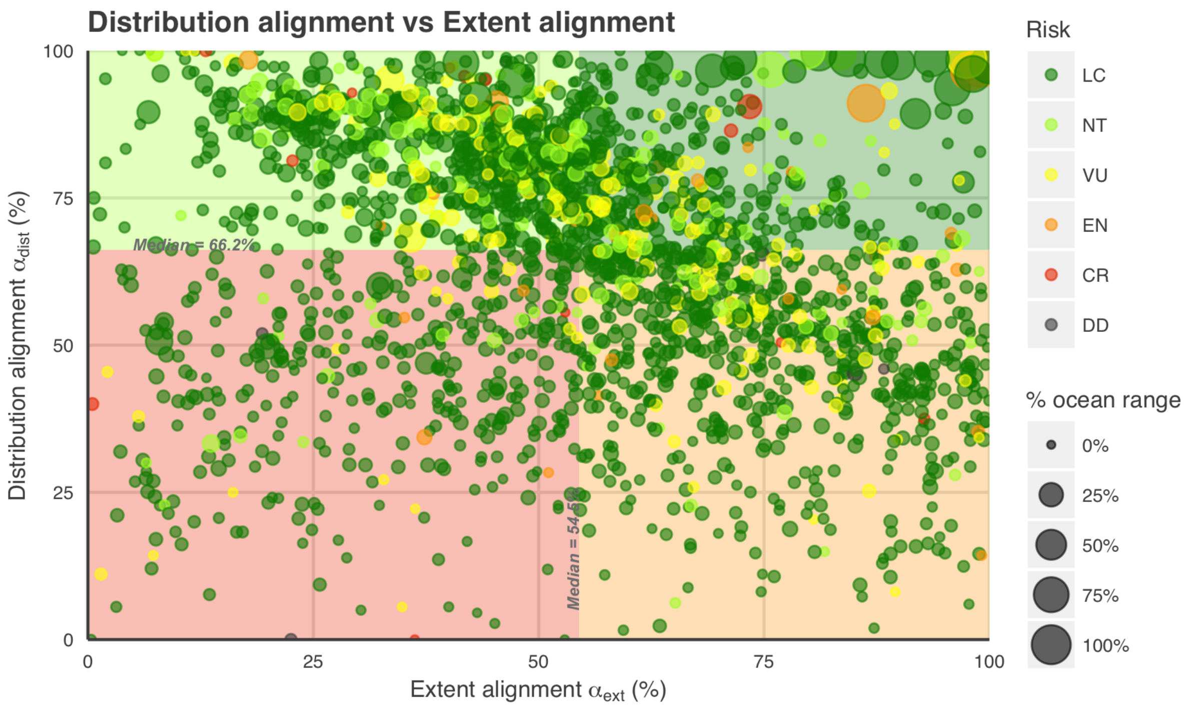


Figure 3 (a): For each paired-map species, we calculated two dimensions of spatial alignment: *distribution alignment*, which we defined as the proportion of the smaller range intersecting the larger range (where on the map); and *extent alignment*, which we defined as the ratio of the smaller range 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 %).

Dividing the map-paired species into quadrants based on median values for each dimension, we can examine the *implications* of four different qualities of alignment. (FIG 3a)

Examining spatial alignment by taxonomic group (FIG 3b), 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. *explain reasoning?*

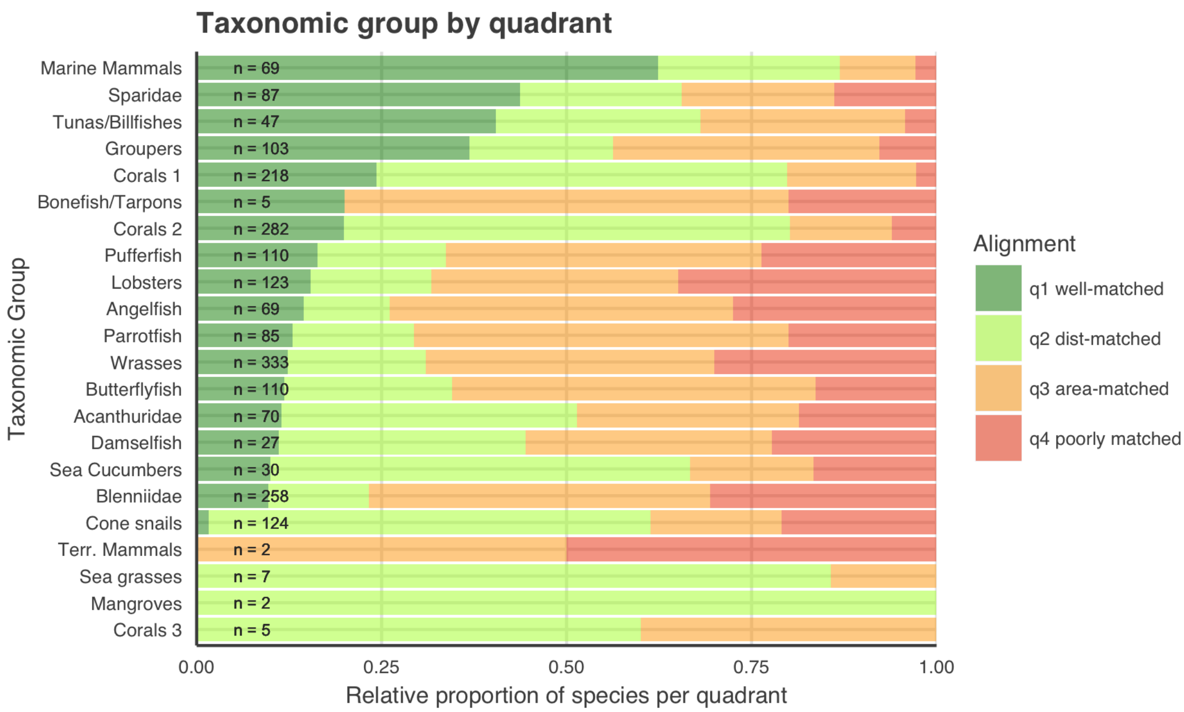


Figure 3 (b): Spatial alignment of paired-map species by taxonomic group.

Quadrant 1 contains species whose map pairs agree in both spatial distribution and the extent of described ranges, as we would expect for well-understood species. Excellent and valuable, but not particularly interesting. The extent-misaligned maps contained in quadrant 2, however, 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 this quadrant (92.4%), the IUCN extent is larger than the AquaMaps extent. By itself, this is not surprising; but 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). We suspect that many of these extent-misaligned map pairs can be easily explained: most corals and reef-associated organisms prefer shallower waters; seafloor depth is explicitly modeled in AquaMaps, but not explicitly considered in IUCN range considerations. ***include recommendation for IUCN range maps to be clipped to bathymetry where appropriate?***

wrap up something with quadrants 3 and 4: *In several instances, the AquaMaps map conformed to a small, local subset of the IUCN map (see SOM for examples - what was the Q4 graph?), in some cases due to some point locality observations being rejected by AquaMaps experts (EXAMPLES) and in other cases due to differences in species identification.* ***bleh - this needs help. Evidence? examples?***

Poorly-aligned rangemaps, regardless of taxon or extinction risk category, indicate species that could benefit from further expert study. ***is this just 'duh'? is there something more important I can say?***

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

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 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. But

---

4.2 Application to MPA Gap Analysis

do the analysis...

Predictions: What happens when using a 0% threshold instead of a 50%? What happens when using IUCN rather than AM? 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?

---

5 Methods

5.0.1 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.

5.0.2 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: 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 4

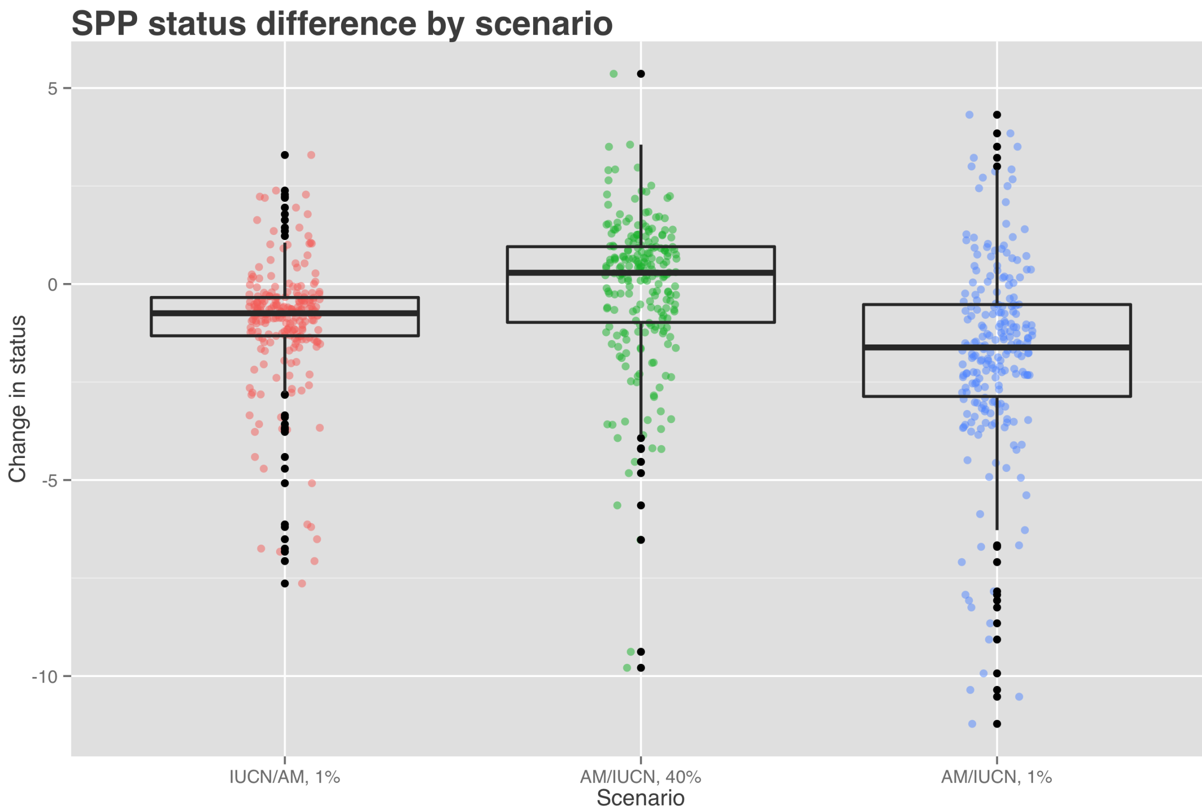


Figure [X] shows the change in status score for the Species Subgoal within the global Ocean Health Index under three different scenarios.

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?