

Aligning marine species range data to better serve science and conservation

Casey C. O'Hara¹, Jamie C. Afflerbach¹, Courtney Scarborough¹, Benjamin S. Halpern^{1,2,3}

1. National Center for Ecological Analysis and Synthesis, University of California, 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,
University of California, 735 State Street, Suite 300, Santa Barbara CA 93101; (805) 892-2500;
ohara@nceas.ucsb.edu

Abstract

Species distribution data provide the foundation for a wide range of ecological research studies and conservation management decisions. Two major efforts to provide marine species distributions at a global scale are the International Union for Conservation of Nature (IUCN), which provides expert-generated range maps that outline the complete extent of a species' distribution; and AquaMaps, which provides model-generated species distribution maps that predict areas occupied by the species. Together these databases represent 24,637 species (92.9% within AquaMaps, 16.3% within IUCN), with only 2,279 shared species. Differences in intent and methodology can result in very different predictions of species distributions, which bear important implications for scientists and decision makers who rely upon these datasets when conducting research or informing conservation policy and management actions. Comparing distributions for the small subset of species with maps in both datasets, we highlight several key examples in which introduced errors drive differences in predicted species ranges. In particular, we find that IUCN maps greatly overpredict coral presence into unsuitably deep waters, and we demonstrate that AquaMaps methodology can produce odd discontinuities at the extremes of a species range. We illustrate the scientific and management implications of these tradeoffs by repeating a global analysis of gaps in coverage of marine protected areas, and find significantly different results depending on how the two datasets are used. Efforts to understand tradeoffs between the two datasets, and ideally to enable the use of both when doing research that requires information on species distributions, will greatly improve the science and policy recommendations around understanding, managing, and protecting marine biodiversity.

36 **Introduction**

37 Knowing where species exist and thrive is fundamental to the sciences of ecology, biogeography,
38 and conservation. This body of science has resulted in various compiled databases of species
39 distribution and range maps which provide foundational information for understanding species
40 diversity and extinction risk, predicting species responses to human impacts and climate change,
41 and managing and protecting species effectively. No species range dataset can claim to represent
42 the "truth" of any species' spatial distribution, but rather each offers its own distinct
43 understanding of that latent distribution. These varying predictions of species presence and
44 absence, driven by intent and methodology, can result in very different predictions of species'
45 ranges. However, conservation managers and policy makers must base their actions on
46 conclusions drawn from these imperfect datasets. It is critical to understand the differences in
47 spatial range datasets and the implications of these differences for conservation research and
48 decision-making.

49 A broad literature exists on methods and results for predicting individual species distributions,
50 but relatively few have been applied to broad suites of taxa globally, particularly in marine
51 contexts. The two most comprehensive, widely-used global-scale repositories that predict marine
52 species ranges throughout the world's oceans are AquaMaps, which rely primarily on model
53 predictions to communicate the distribution of a species based on habitat suitability [1], and
54 range data from the International Union for Conservation of Nature (IUCN), which rely
55 primarily on expert opinion to communicate range size as a criterion for extinction risk
56 assessments [2].

While the two datasets ostensibly describe the same information, i.e., where can a particular species be found, they communicate fundamentally distinct concepts. The IUCN range map for a given species describes the "extent of occurrence" for that species, a contiguous region that encloses all known occurrences and connecting regions, with the explicit caveat that this "does not mean that [the species] is distributed equally within that polygon or occurs everywhere within that polygon" [2]. The AquaMaps map for a species, on the other hand, excludes unsuitable habitats and thus seems to better be defined as "area of occupancy": the subset of the extent of occurrence in which the species is actually likely to be found [1].

The fundamental differences between the datasets suggest that the choice of one over the other should be carefully matched to the purpose for which it is to be used [3], and yet these datasets have been used in hundreds of studies and applications for a wide range of purposes, typically without directly addressing the limitations of the chosen dataset or comparing outcomes between the two. Example uses of the data include assessing marine species status [4–6], evaluating global biodiversity patterns [7–10], predicting species range shifts [11], and setting conservation priorities [12].

We recognize that each dataset provides distinct value for conservation research, and so here we focus not on how the two datasets should ideally be used, but instead, on hidden assumptions and sources of error within each dataset. Importantly, we test the implications of choosing one over the other, and use overlapping species mapped within both datasets to better understand the causes of differences in results from using one or the other. We also repeat a recent analysis of gaps in protection afforded by marine protected areas (MPAs) to examine the implications of these differences for global conservation and management priorities. Because these datasets are

so widely used, it is crucial to understand and acknowledge, and where possible address, data limitations.

Methods and Analysis

About the datasets

The IUCN has published species range maps developed by species experts for 4,027 unique marine species. Experts outline spatial boundaries that define the "limits of distribution" of a given species, based on observation records and informed by expert understanding of species' range and habitat preferences (Fig S1A). IUCN guidelines recommend that boundaries be drawn as a "minimum convex polygon", i.e. "the smallest polygon in which no internal angle exceeds 180 degrees and which contains all the sites of occurrence" [2].

In contrast, AquaMaps has modeled species distribution for 22,889 species based on envelopes of environmental preference, such as temperature, depth, and salinity. These preference envelopes 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 global 0.5° grid to determine suitable habitat, resulting in a "probability of occurrence" for each species (Fig S1B). To roughly constrain species ranges to appropriate georegions, the AquaMaps model uses Food and Agriculture Organization of the United Nations (FAO) Major Fishing Area [13] boundaries. Of the resulting maps, 1,296 (5.7% of the full dataset) have been further refined through an expert review process [14,15].

In total, the two datasets provide range maps for 24,637 species, with a small subset of species mapped in both datasets. For the purposes of this analysis, we elected not to use IUCN data for bird species, which are available separately through Bird Life International [16].

Comparison of taxonomic and regional distribution: To examine the overall taxonomic distribution across the spatial datasets, we grouped species by taxonomic class and data source, and determined the proportion of each class represented in each dataset. 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 (Fig S2). We determined species presence within a grid cell as any non-zero overlap of a species polygon with the cell (Fig S1C). For the AquaMaps dataset, we determined per-cell species count by including all species with non-zero probability of occurrence to best approximate the "extent of occurrence" generally indicated by IUCN maps (Fig S1D).

Comparison of paired maps: Although relatively small in number, overlap species present a unique opportunity to evaluate the two datasets overall. For the species included in both datasets, we examine how well the maps align in both spatial distribution and overall area. Based upon the definitions of extent of occurrence and area of occupancy, we expect that for a given species, the AquaMaps predicted distribution will fall within, and describe a smaller range than, the IUCN predicted distribution. Where these expectations seem to fail, we explore methodological issues that can introduce errors.

We identified "paired map" species using genus and species binomials as a matching key. We used the taxize package [17] in R [18] to standardize species names and synonyms; for species with separate subpopulation maps in IUCN, we combined all subpopulations to create a single

global population. For each of these paired map species, we determined species presence within each spatial cell for each dataset using the same criteria outlined above.

Overlaying paired distribution maps for each species, we defined and calculated *distribution alignment* α_{dist} and *area ratio* α_{area} :

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

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

For each paired map species, A_{small} and A_{large} indicate the smaller and larger range representation (regardless of which dataset) in km^2 . $A_{small \cap large}$ represents the amount of overlapping area between the two datasets. Distribution alignment uses overlapping predictions of presence as a measure of concurrence between the two datasets. Area ratio compares range size, used by IUCN as a criterion to help define extinction risk; it also provides an indicator for frequency of errors of commission (false indication of presence) or omission (false indication of absence).

Examining issues in paired map alignment: Given the wide variation in the alignment of predicted range for the paired map species, we examined two potential drivers of error, one for each dataset. For IUCN data, we explored the assumed consideration of depth as a criterion for range predictions of coral species, a large portion (14%) of the IUCN dataset and an intensely studied taxonomic group whose importance in supporting biodiversity is undisputed. Extracting data from the IUCN API [2] we identified depth limitations of each coral species mapped in the IUCN dataset to verify that none is found below the photosynthetic depth limit of 200 m. Using

the same methodology as shown in Fig S2, we created a 200 m bathymetry raster from a bathymetry spatial dataset (public domain; available from Natural Earth, www.naturalearthdata.com) and masked our IUCN coral rasters to identify mapped coral presence below 200 m. The resulting maps were again compared to the AquaMaps ranges to examine distribution alignment and area ratio.

For AquaMaps data, whose reliance upon FAO Major Fishing Areas [13] to constrain species ranges seems likely to result in abrupt discontinuities in species range where ecologically suitable habitat intersects with human-defined borders, we identified FAO boundaries where such discontinuities are likely to occur. Cropping species distributions to narrow latitudinal bands, we determined species whose eastern or western range limit coincided exactly with a defined FAO boundary. As an illustrative example, we cropped AquaMaps range maps for all Pacific and Indian Ocean species to a band between between 25° S and 20° N latitudes, then identified all species whose eastern range limit within this band coincided with 175° W longitude, the vertical boundary between two prominent FAO regions.

Methods for MPA Gap Analysis case study

To assess the effectiveness of MPAs in protecting biodiversity, Klein et al. [12] compared the coverage of the global MPA network presented by the World Database on Protected Areas (WDPA) [19] to the species ranges described in the 2014 AquaMaps dataset [1]. For the primary analysis, the researchers defined species presence as 50% or greater probability of occurrence.

To reconstruct the primary analysis, we selected the subset of protected areas from the 2014 WDPA dataset classified as IUCN protected area management categories I-IV and spatially overlapping a marine area. The WDPA polygons and marine polygons were rasterized to 0.01°

and then aggregated to AquaMaps native 0.5° cells, to calculate proportion of marine protected area within each cell. After verifying our method using the 2014 AquaMaps dataset, we updated the analysis to use the 2015 AquaMaps dataset, at a presence threshold of 50% (to compare to Klein et al. directly) and 0% (to better compare with IUCN spatial data). To analyze MPA coverage against IUCN spatial data, we extracted IUCN polygon weights per 0.5° cell for each species and compared against the protected area raster.

All processing was completed using R statistical software [18], and all code and intermediate data are available on GitHub at <https://github.com/OHI-Science/IUCN-AquaMaps>.

Results and Discussion

In comparing the IUCN and AquaMaps datasets, it is again important to emphasize that the two differ in both methodology and intent. As area of occupancy is generally defined as a subset of a species' extent of occurrence, AquaMaps ranges (effectively area of occupancy) should generally fall within the bounds of IUCN ranges (effectively extent of occurrence), capturing a smaller total area. However, AquaMaps range maps are created independently from IUCN data and therefore exceptions are certain to arise. Here we are looking for systematic deviations from our expectations that might highlight implications of data use decisions.

Taxonomic and geographic coverage

The two datasets have notably different taxonomic (Fig 1A) and regional (Figs 1B, 1C) coverage. AquaMaps encompasses a broader range of taxa than IUCN, as IUCN spatial data are only available for select taxonomic groups that have been comprehensively assessed. Of the 24,637 species mapped within these datasets, only 2,279 (9.3%) are mapped within both, with

many taxa completely unrepresented in one dataset or the other. While species numbers in both datasets peak in tropical latitudes near the equator, species counts for IUCN maps drop quickly beyond 30°N and 30°S, while AquaMaps includes distribution of species well into temperate latitudes. Together, the limitations of spatial and taxonomic coverage are likely to drive a researcher's choice of dataset far more strongly than the quality, format, or intended purpose of the dataset.

Fig 1. Taxonomic and geographic coverage of AquaMaps and IUCN range data. (A)

Number and proportion of species by taxa included in each dataset. Overlapping species are dominated by bony fishes (983 species, primarily tropical taxa) and corals (396 species). (B, C) Global marine species count per 0.5° cell according to (B) AquaMaps and (C) IUCN. The margin frequency plots show relative species count per cell at each latitude and longitude.

Distribution and range size alignment

Comparing distribution alignment and area ratio for the 2,279 paired map species (Fig 2A), a weak negative linear pattern appears to emerge, suggesting that increasing similarity in range area correlates very slightly with decreasing distribution alignment ($R^2 = .016$). The pattern itself is not particularly important, and emerges simply due to the nature of the analysis and the datasets. In particular, the AquaMaps model tends to extrapolate species ranges into suitable areas beyond known occurrences more frequently than IUCN maps, such that each additional unit of range predicted by AquaMaps will fall in different locations than an additional unit of range predicted using IUCN methodology. For species with dissimilar range areas, predicted distribution for the smaller range can more easily fall within the generous bounds of the larger

range. For species with increasingly similar range areas, differences in methodology become more difficult to "hide," and the distribution alignment generally becomes slightly poorer.

The mean distribution alignment for species included in both datasets was 63%; the mean area alignment was 54.5%. By dividing the paired map species into quadrants based on these means, we highlight categories of relationships to identify patterns in alignment differences.

Representative maps from each category are provided in the supporting materials (Fig S3).

Fig 2. Comparison of alignment between AquaMaps and IUCN range data. (A)

Distribution alignment (overlap of smaller range within larger) versus area ratio (the ratio of smaller range area to the larger range area) for 2,279 species included in both IUCN and AquaMaps datasets. The upper right quadrant comprises species whose maps largely agree in both spatial distribution and the extent of described ranges ($n = 527$; 22.8 %). The upper left quadrant comprises species whose maps agree well in distribution, but disagree in area ($n = 709$; 30.6 %). The lower right quadrant includes species for which the paired maps generally agree in range area, but disagree on where those ranges occur ($n = 635$; 27.4 %). The lower left quadrant indicates species for which the map pairs agree poorly in both area and distribution ($n = 443$; 19.1 %). (B) Alignment quadrant breakdown of species by taxonomic group.

The upper right quadrant includes the species ($n = 527$) whose described ranges are above average in alignment of both spatial distribution and area. 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 their extent of occurrence and area of occupancy both span nearly the entire map.

225 The area-mismatched ranges contained in the upper left quadrant ($n = 709$) include many species
226 whose spatial distribution is similar, but where one range is notably larger than the other. For
227 88% of the species in this quadrant, the IUCN range is larger than the AquaMaps range, as
228 expected when comparing an extent of occurrence to an area of occupancy map. However, the
229 low area ratio suggests a high rate of commission and/or omission errors by one or both datasets.
230 Below we explore one underlying source of commission errors.

231 Species found in the lower right quadrant ($n = 635$) represent cases of "two wrongs make a
232 right." For these species, IUCN and AquaMaps both predict ranges extending far beyond the
233 overlapping region, but the methodological differences result in very different extrapolations.
234 Consequently, area ratios are high, though the poor distribution alignment indicates that one or
235 both datasets are introducing significant errors. In this quadrant, the IUCN range is the larger for
236 only 56.5% of species, defying our expectations based on definitions of area of occupancy and
237 extent of occurrence.

238 The lower left quadrant includes species ($n = 443$) where alignment is poor in both dimensions.
239 In this quadrant, the IUCN range is larger for only 24.6% of species; this again defies our
240 expectations for extent of occurrence and area of occupancy range predictions. Data-poor species
241 are more common in this quadrant; indeed, the median number of species occurrence records
242 (averaging occurrences from the Ocean Biogeographic Information System (OBIS) [20] and the
243 Global Biodiversity Information Facility (GBIF) [21]) for this quadrant is 24 records, compared
244 to a median of 97 records for species across the other three quadrants. The AquaMaps dataset
245 offers its own quality metric based on the number of unique 0.5° cells containing valid
246 occurrences; for this quadrant, the median "occurcells" is 11 compared to a median of 40 across
247 the other three quadrants. Care should be taken when using distribution and range maps based

upon fewer observations, as they clearly bear greater uncertainty; AquaMaps explicitly warns against using any of its maps generated with an "occurcells" count fewer than 10 [1].

Coral depth exploration

Because corals dominate the upper-left "distribution-aligned" quadrant of Fig 2A (n = 237; 33.4% of all species in this quadrant), we explored implications of explicitly restricting IUCN ranges to depths based on species' life histories. This adjustment was not necessary for AquaMaps data because models explicitly include ocean depth preference as a parameter. While depth is recommended by the IUCN as a criterion for providers of range maps ("The limits of distribution can be determined by using known occurrences of the species, along with the knowledge of habitat preferences, remaining suitable habitat, elevation limited, and other expert knowledge of the species and its range." [2]), it is not presented as a requirement, so we cannot take its inclusion for granted.

Fig 3A shows aggregated ranges of the 463 coral species mapped in the IUCN dataset, with their ranges broken into proportional area deeper and shallower than 200 m. According to IUCN descriptions, none of these species is indicated to occur deeper than 200 m, and 94% are confined to waters shallower than 50 m; seven of the mapped species had no reported depth information. Clipping coral ranges to shallower than 200 m eliminated an average of 47.6% of the total predicted area while still allowing for a generous estimate of suitable habitat.

Fig 3. Effect of 200 m depth constraint on IUCN range maps for coral species. (A)

Aggregate map combining ranges of the 463 coral species mapped in the IUCN dataset, showing raw ranges and ranges clipped to 200 m depth. (B) Alignment quadrant breakdown

of paired map coral species using original data from IUCN and AquaMaps (as in Fig 2B) and the same species with IUCN ranges clipped to 200 m depth.

In constraining coral ranges to appropriate depths, we see a strong increase in the apparent alignment of species maps between IUCN and AquaMaps (Fig 3B). Membership in the "well-aligned" quadrant jumped from 22.4% to 76.2%, with a corresponding decrease in all other quadrants. By excluding the unsuitable areas from IUCN's extent of occurrence, we eliminate preventable commission errors and more closely approximate the area of occupancy described by AquaMaps. See Fig S4 to examine the shifts of individual species among the quadrants.

The true distribution of each of these corals remains imperfectly known. Certainly some commission errors result from extent of occurrence boundaries including areas where species do not actually exist (as recognized by the definition of this method), and others may be due to experts taking a precautionary (i.e., generous) approach to likely occurrence. Yet a simple and sensible shift in method drastically decreases the likelihood of introducing commission errors, with little chance of introducing omission errors, greatly improving our confidence in the remaining reported distributions for most purposes. This change applies just as readily to the IUCN coral maps that are not included in the paired map analysis, and likely to other reef-associated flora and fauna. While species depth preferences are an easy and consistent means of constraining range predictions, other conditions such as salinity and temperature could be cautiously used to refine the results of expert opinions.

Georegional constraint exploration

From the entirety of the AquaMaps dataset, we identified 3,208 Indo-Pacific species whose equatorial distributions (between 25° S and 20° N) encounter an eastern range limit at 175° W. A

clear discontinuity in species distributions of a single example species (Fig 4A) and all 3,208 species in aggregate (Fig 4B) matches perfectly with FAO region 77 [13]; other discontinuities are apparent at other FAO boundaries, despite these boundaries not being actively studied in this analysis.

Fig 4. Effect of FAO Major Fishing Area constraints on AquaMaps distributions. (A)

AquaMaps species distribution of *Hoplichthys regani*, the ghost flathead, with known occurrence records. (B) Aggregated AquaMaps predicted ranges for 3,208 species whose equatorial distribution encounters an eastern discontinuity exactly at 175° W, the boundary between FAO Major Fishing Areas 71 and 77 (shown in blue). Other FAO area boundaries create additional clear discontinuities.

FAO Major Fishing Area boundaries provide a readily available method to roughly constrain AquaMaps predictions to appropriate ocean basins, thus eliminating a large source of potential commission errors and enabling rapid modeling of thousands of species ranges. However, these boundaries are defined for statistical purposes based on economic and political considerations rather than ecological considerations, and can result in odd discontinuities in species range predictions where otherwise suitable habitat is excluded. While such a discontinuous boundary would likely be obvious when inspecting the distribution of an individual species, the distinction is likely to be obscured when aggregating many species ranges as is typical for biodiversity or conservation studies.

The ratio of the total predicted range for a species to the number of "occurcells" used to generate the map provides a measure of the degree to which AquaMaps extrapolates range area from limited data. For example, AquaMaps predicts a total range of 5.4 million km² for both the round ray *Rajella fyllae* and the brittle star *Ophiothrix plana*; but the map for *R. fyllae* is generated

using 116 "occurcells" (for a data extrapolation rate of 46,800 km² per cell) while the map for *O. plana* is generated using only four (for an extrapolation rate of 1,360,000 km² per cell).

By this measure, the 3,208 species range maps included in Fig 4 tend to extrapolate farther based on limited data: a mean predicted range of 827,200 km² per "occurcell" compared to a mean predicted range of 485,300 km² per "occurcell" for the overall AquaMaps dataset. This suggests that the FAO boundaries may not be sufficient to adequately constrain computer-generated ranges. To reduce the incidence of commission errors due to aggressive extrapolation, it may be desirable to fine-tune the computer model output with additional filters, such as ecoregional constraints, e.g. Marine Ecoregions of the World [22], or distance-based methods, e.g. inverse distance weighting to enforce proximity to known observations. Expert review, though time-consuming, is the most certain route to boosting confidence in these predicted distributions.

Case Study: MPA Gap Analysis

Klein et al. [12] compare the global distribution of species to the global distribution of marine protected areas to assess how well current MPAs overlap with species ranges and identify which species fall through gaps in protection. The study relied on the 2014 version of the AquaMaps database, using a probability of occurrence threshold of 50% or greater, to determine species presence, and the World Database of Protected Areas to define zones of marine protection. They found that the global MPA network leaves 90.5% of marine species with less than 5% of their overall range represented within MPAs, and 1.4% of species have no protection at all (i.e., "gap" species).

We recalculated the amount of under-protected and gap species using all available IUCN species ranges, as well as the 2015 AquaMaps data at a 50% threshold to replicate the original methods

and a 0% threshold to more closely approximate the extent of occurrence represented by IUCN data (Fig 5). Comparing the IUCN results to the AquaMaps 2015 results (at 0% threshold) we found a five-fold increase in the proportion of gap species (6.4% of species vs. 1.2%) and dramatically larger proportion of species with less than 2% of their range protected (73.2% of species vs. 47.7%). However, this comparison also indicates a larger proportion of well-protected species with greater than 10% of range protected (2.9% of species vs. 1.5%).

Fig 5. MPA gap analysis results based upon alternate choices of datasets. Percent of species range covered by MPAs based upon methods in Klein et al. (2015). Scenario 1 replicates the original results, measuring protected range of species in AquaMaps 2014 dataset, with a 50% presence threshold, against the 2014 World Database of Protected Areas, filtered for IUCN categories I-IV that overlap marine areas. Scenario 2 updates the results using AquaMaps 2015, showing very small changes despite the inclusion of an additional 5,545 species. Scenario 3, using 2015 AquaMaps data, drops the presence threshold to zero, showing an expected decrease in gap species, but also a decrease in species with 5% or greater protected range. Scenario 4 examines species MPA coverage using only the IUCN dataset.

To achieve more comprehensive taxonomic coverage of species ranges it may be desirable to use these two datasets together. The Ocean Health Index (OHI) is one analysis that uses both IUCN and AquaMaps spatial data, in conjunction with extinction risk categories from the IUCN Red List of Threatened Species, to calculate area-weighted extinction risks for the Species component of the index. To combine these data sets, OHI uses all available IUCN range maps, and then supplements with AquaMaps data (with "presence" determined by 40% or greater probability of occurrence) for species whose ranges are not represented within IUCN; species without valid

Red List extinction risk categories are excluded from analysis [4,5]. Such a combination of two inherently different datasets increases the number of species represented within the analysis, but the differences in range representations between the datasets is likely to distort the results.

While it may be unrealistic to "fix" one data set to match the other, we may be able to reduce the impact of the tradeoffs inherent in each. Trimming unsuitable habitat from the IUCN's extent of occurrence maps, for example by explicitly clipping them to appropriate depths for corals and reef-associated species, would result in maps that more closely approximate areas of occupancy. Conversely, including all AquaMaps cells with a non-zero "probability of occurrence" (rather than using a probability threshold to determine presence, e.g. greater than 40% for the Ocean Health Index [4,5] or 50% for the MPA gap analysis [12]) would allow for the most generous inclusion of species range, resulting in maps that more closely, though still imperfectly, approach the IUCN's extents of occurrence.

Conclusions

No dataset can ever claim to know the "truth" of the location and extent of marine biodiversity. AquaMaps and IUCN range maps show strong agreement for many well-studied species, but for many others, substantial differences arise from differences in methodology and intent of each dataset. While the decision of which dataset to use should ideally be driven by the intended purpose for which it was created, the fact is that geographic and taxonomic coverage will likely be a more important factor in determining which dataset is used. Recognizing and acknowledging the advantages and differences of the range maps presented by these datasets will increase their utility for research and conservation actions. Conclusions drawn from each of these datasets could paint dramatically different pictures of global marine biodiversity or the

effectiveness of conservation management decisions. By highlighting the distinctions between these two important and fundamentally different marine species range datasets, we improve our ability to inform strategic and effective conservation policy that supports a resilient ocean ecosystem.

Acknowledgments

We are thankful to Melanie Frazier, Julia Stewart Lowndes, Halley Froelich, and Claire Runge for their insightful comments on this manuscript, and to the National Center for Ecological Analysis and Synthesis for computation support. We acknowledge financial support from the Gordon and Betty Moore Foundation.

References

1. Kaschner K, Rius-Barile J, Kesner-Reyes K, Garilao C, Kullander S, Rees T, et al. AquaMaps: Predicted range maps for aquatic species [Internet]. 2015. Available: www.aquamaps.org
2. IUCN. The IUCN Red List of Threatened Species [Internet]. International Union for the Conservation of Nature. 2015. Available: <http://www.iucnredlist.org>
3. Gaston KJ, Fuller RA. The sizes of species' geographic ranges. *Journal of Applied Ecology*. 2009;46: 1–9. doi:[10.1111/j.1365-2664.2008.01596.x](https://doi.org/10.1111/j.1365-2664.2008.01596.x)
4. Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, Katona SK, et al. An index to assess the health and benefits of the global ocean. *Nature*. 2012;488: 615–620. doi:[10.1038/nature11397](https://doi.org/10.1038/nature11397)
5. Halpern BS, Longo C, Lowndes JSS, Best BD, Frazier M, Katona SK, et al. Patterns and Emerging Trends in Global Ocean Health. Tsikliras AC, editor. *PLOS ONE*. 2015;10: e0117863. doi:[10.1371/journal.pone.0117863](https://doi.org/10.1371/journal.pone.0117863)
6. Selig ER, Longo C, Halpern BS, Best BD, Hardy D, Elfes CT, et al. Assessing Global Marine Biodiversity Status within a Coupled Socio-Ecological Perspective. Guichard F, editor. *PLoS ONE*. 2013;8: e60284. doi:[10.1371/journal.pone.0060284](https://doi.org/10.1371/journal.pone.0060284)
7. Coll M, Piroddi C, Steenbeek J, Kaschner K, Ben Rais Lasram F, Aguzzi J, et al. The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. Bograd SJ, editor. *PLoS ONE*. 2010;5: e11842. doi:[10.1371/journal.pone.0011842](https://doi.org/10.1371/journal.pone.0011842)

- 411 8. Martin C, Fletcher R, Jones M, Kaschner K, Sullivan E, Tittensor DP, et al. Manual of marine
412 and coastal datasets of biodiversity importance. United Nations Environment Programme; 2014
413 May.
- 414 9. Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, et al. The biodiversity
415 of species and their rates of extinction, distribution, and protection. *Science*. 2014;344: 1246752–
416 1–1246752–10. doi:[10.1126/science.1246752](https://doi.org/10.1126/science.1246752)
- 417 10. Kaschner K, Tittensor DP, Ready J, Gerrodette T, Worm B. Current and Future Patterns of
418 Global Marine Mammal Biodiversity. Bograd SJ, editor. *PLoS ONE*. 2011;6: e19653.
419 doi:[10.1371/journal.pone.0019653](https://doi.org/10.1371/journal.pone.0019653)
- 420 11. García Molinos J, Halpern BS, Schoeman DS, Brown CJ, Kiessling W, Moore PJ, et al.
421 Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate*
422 *Change*. 2015;6: 83–88. doi:[10.1038/nclimate2769](https://doi.org/10.1038/nclimate2769)
- 423 12. Klein CJ, Brown CJ, Halpern BS, Segan DB, McGowan J, Beger M, et al. Shortfalls in the
424 global protected area network at representing marine biodiversity. *Scientific Reports*. 2015;5:
425 17539. doi:[10.1038/srep17539](https://doi.org/10.1038/srep17539)
- 426 13. FAO Fisheries and Aquaculture Department. CWP Handbook of Fishery Statistical
427 Standards. Section H: FISHING AREAS FOR STATISTICAL PURPOSES. Food; Agriculture
428 Organization of the United Nations; 2015.
- 429 14. Kaschner K, Watson R, Trites AW, Pauly D, others. Mapping world-wide distributions of
430 marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology*
431 *Progress Series*. 2006;316: 2–3. Available: <http://www.vliz.be/imisdocs/publications/100462.pdf>

432 15. Ready J, Kaschner K, South AB, Eastwood PD, Rees T, Rius J, et al. Predicting the
433 distributions of marine organisms at the global scale. *Ecological Modelling*. 2010;221: 467–478.
434 doi:[10.1016/j.ecolmodel.2009.10.025](https://doi.org/10.1016/j.ecolmodel.2009.10.025)

435 16. BirdLife International and NatureServe. Bird species distribution maps of the world. BirdLife
436 International, Cambridge, UK; NatureServe, Arlington, USA; 2015.

437 17. Scott Chamberlain, Eduard Szocs. Taxize - taxonomic search and retrieval in R.
438 F1000Research. 2013; Available: <http://f1000research.com/articles/2-191/v2>

439 18. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna,
440 Austria: R Foundation for Statistical Computing; 2016. Available: <https://www.R-project.org>

441 19. IUCN, UNEP-WCMC. The World Database on Protected Areas (WDPA) [Internet].
442 Cambridge, UK: UNEP-WCMC. 2014. Available: www.protectedplanet.net

443 20. OBIS. Data from the Ocean Biogeographic Information System [Internet]. Intergovernmental
444 Oceanographic Commission of UNESCO. 2016. Available: <http://www.iobis.org>

445 21. GBIF. Global Biodiversity Information Facility (GBIF) Memorandum of Understanding
446 [Internet]. Global Biodiversity Information Facility. 2010. Available:
447 <http://www.gbif.org/resource/80661>

448 22. Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson MAX, et al. Marine
449 ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience*. 2007;57:
450 573–583. Available: <https://bioscience.oxfordjournals.org/content/57/7/573.full>

References for supplemental figures

23. www.aquamaps.org, version of Aug. 2013. Reviewed distribution maps for *Thunnus alalunga* (Albacore), with modelled year 2100 native range map based on IPCC A2 emissions scenario. Web. Accessed 11 Sep. 2016.
24. International Union for Conservation of Nature (IUCN) 2011. *Thunnus alalunga*. The IUCN Red List of Threatened Species. Version 2016-2
25. www.aquamaps.org, version of Aug. 2013. Computer generated distribution maps for *Conus episcopatus*, with modelled year 2100 native range map based on IPCC A2 emissions scenario. Web. Accessed 11 Sep. 2016.
26. Conch Books, Hackenheim, Germany 2013. *Conus episcopatus*. The IUCN Red List of Threatened Species. Version 2016-2
27. www.aquamaps.org, version of Aug. 2013. Computer generated distribution maps for *Kajikia albida* (Atlantic white marlin), with modelled year 2100 native range map based on IPCC A2 emissions scenario. Web. Accessed 11 Sep. 2016.
28. International Union for Conservation of Nature (IUCN) 2011. *Kajikia albida*. The IUCN Red List of Threatened Species. Version 2016-2
29. www.aquamaps.org, version of Aug. 2013. Reviewed distribution maps for *Acanthurus nigroris* (Bluelined surgeonfish), with modelled year 2100 native range map based on IPCC A2 emissions scenario. Web. Accessed 11 Sep. 2016.
30. International Union for Conservation of Nature (IUCN) 2012. *Acanthurus nigroris*. The IUCN Red List of Threatened Species. Version 2016-2

472 31. www.aquamaps.org, version of Aug. 2013. Computer generated distribution maps for *Conus*
473 *magnificus*, with modelled year 2100 native range map based on IPCC A2 emissions scenario.
474 Web. Accessed 11 Sep. 2016.

475 32. Conch Books, Hackenheim, Germany 2013. *Conus magnificus*. The IUCN Red List of
476 Threatened Species. Version 2016-2

477

Supporting figure captions

S1 Figure. Examples of AquaMaps and IUCN raw species range data. (A, B) AquaMaps distribution of *Thunnus alalunga* (Albacore Tuna) [23] showing (A) global range with varying probabilities of occurrence assigned to 0.5° grid cells and (B) presence within 0.5° grid cells. (C, D) IUCN distribution of *T. alalunga* [24] represented as (C) extent of occurrence polygons and (D) presence within 0.5° grid cells.

S2 Figure. Rasterizing shapefiles provided by IUCN. A portion of the *T. alalunga* range map [24] is used to exemplify the rasterization process. To enable direct comparison of IUCN species ranges to AquaMaps species ranges, the raw IUCN polygon (A) is overlaid with a 0.5° degree grid matching the AquaMaps grid (B). Each cell is assigned a value of "present" if the cell overlaps any portion of the polygon (C). The resulting raster (D).

S3 Figure. Representative species maps to illustrate each quadrant from Fig 2A. Each map is positioned to match its quadrant in Fig 2A. FAO Major Fishing Area boundaries [11] are outlined in light grey. (A) Distribution-aligned: *Conus episcopatus*, the dignified cone snail [25, 26]. Distributions show excellent overlap in the western Pacific, though IUCN range extends well beyond the bounds of the AquaMaps range. (B) Well-aligned: *Kajikia albida*, the Atlantic white marlin [27, 28]. Distributions from each data set show nearly complete overlap, and very similar range size. (C) Poorly aligned: *Acanthurus nigroris*, the blue-lined surgeonfish [29, 30]. IUCN predicts species distribution only near the Hawaiian islands; AquaMaps predicts extensive distribution throughout the central and western Pacific Ocean. The datasets align in neither distribution nor range size. (D) Area-aligned: *Conus magnificus*, the magnificent cone snail [31, 32]. Distributions overlap in the southern Pacific, but align poorly elsewhere. The range sizes are similar.

S4 Figure. Shift in alignment of paired-map coral species due to clipping IUCN ranges

to areas shallower than 200 m. The grey lines represent the change in apparent alignment

for a single species. Most coral species shift rightward from the upper left quadrant to the

upper right, improving in area alignment with little if any loss in distribution alignment, since

in general, only unsuitable habitat has been removed. Leftward shifts can be seen in species

whose larger original range is represented in AquaMaps; by trimming IUCN ranges, the area

ratio becomes smaller.