

# AquaMaps: Concept, Algorithms and Data Sources for Modelling the Distribution of Marine Organisms

## *AquaMaps for Marine Species*

AquaMaps is a tool for generating model-based, large-scale predictions of natural occurrences of marine species. The model uses estimates of environmental preferences with respect to depth, water temperature, salinity, primary productivity, dissolved oxygen, and association with sea ice or coastal areas. These estimates of species preferences, called environmental envelopes, are derived from large sets of occurrence data available from online collection databases such as GBIF ([gbif.org](http://gbif.org)) and OBIS ([obis.org](http://obis.org)), and from independent knowledge from the literature about the distribution of a given species and its habitat usage that are available in FishBase ([www.fishbase.org](http://www.fishbase.org)) and in SeaLifeBase ([www.sealifebase.org](http://www.sealifebase.org)). The environmental envelopes are matched against local environmental conditions to determine the suitability of a given area in the ocean for a particular species. Predictions of relative probabilities of species occurrence are shown as color-coded species range maps in a global grid of half-degree latitude and longitude cell dimensions. The maps are displayed on the web through the use of [C-squares Mapper](#) developed at CSIRO Marine and Atmospheric Research in Australia (Rees, 2002, 2003).

The AquaMaps approach of incorporating species occurrences and expert knowledge into an environmental envelope is modified from an ecological niche model originally developed for predicting global distributions of marine mammals (Kaschner et al., 2006). It is specifically applied to correct for biases in occurrence data such as non-representative coverage of a species' distribution, biases in sampling effort and data provision, and species misidentifications. Further, this approach is applicable to a wide range of marine organisms thus allowing AquaMaps modeling of both fish and non-fish species.

AquaMaps predictions have been validated using independent and effort-corrected survey data (Ready et al., 2010). The performance of the model is found to compare well with other presence-only species distribution models such as GARP - Genetic Algorithm for Rule Set Production (Anderson et al., 2003), MaxEnt - Maximum Entropy Modeling (Phillips et al., 2006), GLMs - generalized linear models and GAMs - generalized additive models (McCullagh and Nelder, 1989; Hastie, 1991).

## HOW MARINE AQUAMAPS WORKS

AquaMaps predictions of species distributions are generated in a two-step process. In the first step, maps are computer-generated using algorithm-derived input parameter settings based on occurrence data filtered with information on the distribution and habitat usage of a species (e.g., depth range, geographic range limits, environment occupied according to adult feeding or breeding behavior). In the second step, experts can review, edit and approve the computer-generated maps. These reviewed maps can only be updated by experts. In contrast, the computer-generated maps are rerun every 1-2 years, as new species data become available.

## *Computer-generated AquaMaps*

### Input Data

The process of generating a modelled species distribution map begins with the consolidation of the following data:

**Occurrence Data:** These consist of geo-referenced occurrence records harvested from GBIF (Global Biodiversity Information Facility) and OBIS (Ocean Biodiversity Information System), as well as from museum collections and the literature available in FishBase and SeaLifeBase. Since 2010, we also use “country points” as proxy occurrence data based on a species’ reported presence in a country. These consist of two sets of geographic coordinates assigned to a country, one in shallow and another in deep water. Either point or both are applied depending on the depth range of a species. The country points are meant to boost occurrence data needed to model the distribution of a species. Point data gathered for a species are used to identify half-degree cells in an authority file from which environmental parameter values will be extracted to generate the species environmental envelope.

**Bounding Box:** This consists of manually encoded geographic coordinates that define the limits of the distributional range of a species. Recorded are the northernmost, southernmost, westernmost and easternmost limits of a species’ known native range, based on published maps or from the literature. The bounding box is used to identify “good cells”, i.e., half-degree cells that contain occurrence data located within the bounding box. This is done to prevent processing of “bad cells” resulting from erroneous or doubtful species occurrence data. The bounding box is also used to prevent the native range of the species from showing in areas that contain suitable habitat but are not occupied by the species. For example, a species that is endemic to the Atlantic will not be shown in the Pacific even though it would find suitable habitat there.

**FAO Areas:** These are large geographical marine fishing areas designated by FAO (Food and Agriculture Organization of the United Nations) for statistical purposes. All species in FishBase and SeaLifeBase are assigned to these areas. For AquaMaps, only those FAO areas where a species is endemic or native are considered. Similar to the bounding box, FAO areas are used to identify “good” and “bad” cells. A minimum of 10 different “good” cells is recommended to generate the species envelope and map for a species.

**Depth Ranges:** This information is encoded in FishBase and SeaLifeBase and consists of the minimum and maximum values of the absolute (min–max) and the common depth ranges of a species, based on the literature. These values, in meters, make up the depth envelope of a species. In the absence of published values, depth range estimates are used (see ESTIMATEDDepth table). These estimates are based on other information related to the species such as habitat description, depth range of congeners or other species of the same family, or minimum and/or maximum depths recorded from survey data. Depth data are used to limit predicted distributions to areas consistent with the depth range of a species. They are also used to determine whether surface or bottom temperature and salinity are to be used when modelling the distribution of a given species.

**Pelagic Flag:** This flag indicates whether the temperature and salinity preferences of a species are based on surface or bottom values. Entries in the SPECIES table regarding the environment for adult feeding and breeding behavior are converted to True or False statements (see Table 1). If True, the species is in the water column well above and independent of the bottom, such as pelagic-oceanic and bathypelagic species. If False, species occurrence is influenced by bottom depth and will thus require the application of depth filters when computing probability of occurrence from the depth envelope.

**Table 1. Pelagic flag equivalent of species environment entries in the FishBase and SeaLifeBase SPECIES table.**

<b>Environment/Province</b>	<b>Pelagic Flag</b>
<b>Pelagic-neritic</b>	<b>False</b>
<b>Pelagic-oceanic</b>	<b>True</b>
<b>Demersal</b>	<b>False</b>
<b>Benthopelagic</b>	<b>False</b>
<b>Reef-associated</b>	<b>False</b>
<b>Bathypelagic</b>	<b>True</b>
<b>Bathydemersal</b>	<b>False</b>

#### Environmental Factors

The AquaMaps algorithm for marine species uses seven environmental layers representing key physical and biological factors that structure the distribution range of many species at large scales and are thus used as predictors of species presence. These include depth, sea temperature, salinity, primary productivity, dissolved oxygen, sea ice concentration and distance to land. Geo-referenced values for these factors are provided at 0.5° resolution and are stored as sets of cell attributes in a Half-degree Cell Authority File (HCAF) along with their associated LOICZ (<http://loicz.org/loicz/>) and C-squares ID numbers ([www.marine.csiro.au/csquares/](http://www.marine.csiro.au/csquares/)). The HCAF contains such attributes for a grid of 179,904 half-degree cells over marine waters. Each cell contains corresponding values from the environmental layers listed in Table 2 below, covering present-day and future ocean conditions.

**Tables 2. Environmental layers used by the marine AquaMaps species distribution model**

Environmental Factor	Variables	Description	Unit	Period	Range	Source
<b>Depth</b>	DepthMin	Minimum, maximum and mean cell bathymetry	m	Unchanged from present through 2050	0 to 7113	Pauly et al. (2020)
	DepthMax	derived from ETOPO 2min negative bathymetry			0 to 10654	
	DepthMean	elevation			0 to 8672	
<b>Sea temperature</b>	SSTAnMean	Mean surface and bottom sea temperature at present-day condition <sup>a</sup>	°C	Present	-1.80 to 30.14	Tyberghein et al. (2012); Assis et al. (2017)
	SBTAnMean				-1.84 to 31.14	
	SST2050	Magnitude of change in mean surface/bottom sea temperatures by 2050; difference between modelled 2050 and modelled 2007 mean surface/bottom temperature (MPIM-ESM model, RCP 8.5 scenario) <sup>b</sup>	°C	-	-1.22 to 4.34	Giorgetta et al. (2013)
	SBT2050				-1.66 to 4.32	
<b>Salinity</b>	SST2050c	Debiased 2050 mean surface/bottom sea temperature; sum of present-day mean sea temperature and magnitude of change in mean sea temperature by 2050 (e.g., SST2050c=SSTAnMean + SST2050)	°C	2050 (RCP 8.5)	-1.79 to 31.46	Kesner-Reyes et al. (2019)
	SBT2050c				-1.84 to 32.02	
	SalinityMean	Mean surface and bottom salinity at present-day condition <sup>a</sup>	PSU	Present	0.06 to 40.59	Tyberghein et al. (2012); Assis et al. (2017)
	SalinityBMean				4.99 to 40.73	
<b>Primary productivity</b>	Salinity2050	Magnitude of change in mean surface/bottom salinity by 2050; difference between modelled 2050 and modelled 2007 mean surface/bottom salinity (MPIM-ESM model, RCP 8.5 scenario) <sup>b</sup>	PSU	-	-2.97 to 2.68	Giorgetta et al. (2013)
	SalinityB2050				-2.97 to 2.68	
	Salinity2050c	Debiased 2050 mean surface/bottom salinity; sum of present-day mean salinity and magnitude of change in mean salinity by 2050 (e.g., Salinity2050c=SalinityMean + Salinity2050)	PSU	2050 (RCP 8.5)	0.06 to 40.60	Kesner-Reyes et al. (2019)
	SalinityB2050c				5.05 to 40.78	
<b>Primary productivity</b>	PrimProdMean	Proportion of mean primary productivity at present-day condition <sup>a</sup>	mgC·m <sup>-3</sup> ·d <sup>-1</sup>	Present	0.09 to 256.71	Tyberghein et al. (2012); Assis et al. (2017)
	PrimProd2050	Magnitude of change in mean primary productivity by 2050; difference between modelled 2050 and modelled 2007 mean primary productivity (MPIM-ESM model, RCP 8.5 scenario) <sup>b</sup>	mgC·m <sup>-3</sup> ·d <sup>-1</sup>	-	-6.29 to 5.21	
	PrimProd2050c	Debiased 2050 mean primary productivity; sum of present-day mean primary productivity and magnitude of change in mean primary productivity by	mgC·m <sup>-3</sup> ·d <sup>-1</sup>	2050 (RCP 8.5)	0 to 256.78	Kesner-Reyes et al. (2019)

		2050 (e.g., $\text{PrimProd2050c} = \text{PrimProdMean} + \text{PrimProd2050}$ )				
<b>Dissolved molecular oxygen<sup>c</sup></b>	OxyBMean	Mean bottom dissolved molecular oxygen at present-day condition <sup>a</sup>	mmol <sup>-3</sup>	Present	0 to 413.87	Tyberghein et al. (2012); Assis et al. (2017)
	OxyB2050	Magnitude of change in mean bottom dissolved molecular oxygen by 2050; difference between modelled 2050 and modelled 2007 mean bottom dissolved molecular oxygen (MPIM-ESM model, RCP 8.5 scenario) <sup>b</sup>	mmol <sup>-3</sup>	-	-56.14 to 57.20	Giorgetta et al. (2013)
	OxyB2050c	Debiased 2050 mean bottom dissolved molecular oxygen; sum of present-day mean bottom dissolved molecular oxygen and magnitude of change in mean dissolved molecular oxygen by 2050 (e.g., $\text{OxyB2050c} = \text{OxyBMean} + \text{OxyB2050}$ )	mmol <sup>-3</sup>	2050 (RCP 8.5)	0 to 412.32	Kesner-Reyes et al. (2019)
<b>Sea ice concentration</b>	IceConAnn	Mean percent sea ice concentration (area proportion) at present-day condition <sup>a</sup>	as decimal	Present	0 to 0.97	Tyberghein et al. (2012); Assis et al. (2017)
	IceCon2050	Magnitude of change in mean percent sea ice concentration by 2050; difference between modelled 2050 and modelled 2007 mean percent sea ice concentration (MPIM-ESM model, RCP 8.5 scenario) <sup>b</sup>	as decimal	-	-99.99 to 0.03	Giorgetta et al. (2013)
	IceCon2050c	Debiased 2050 mean percent sea ice concentration; sum of present-day mean percent sea ice concentration and magnitude of change in mean percent sea ice concentration by 2050 (e.g., $\text{IceConB2050c} = \text{IceConAnn} + \text{IceCon2050}$ )	as decimal	2050 (RCP 8.5)	0 to 0.92	Kesner-Reyes et al. (2019)
<b>Distance to land</b>	LandDist	Distance of marine cell to nearest coastal cell	km	Unchanged from present through 2050	0 to 2709	Pauly et al. (2020)

<sup>a</sup> Present-day (current) conditions are based on monthly averages for the years 2000-2014 of pre-processed global ocean re-analyses incorporating satellite and in situ observations

<sup>b</sup> Modelled 2050 value is based on monthly averages for the years 2046-2055, while modelled 2007 on monthly averages for the years 2000-2014 (MPIM-ESM-MR model, RCP 8.5 scenario)

<sup>c</sup> Applied to deep sea species only

With regard to 2050 ocean conditions, the AquaMaps model does not make straightforward use of the Max Planck Institute Earth System Model in 'Medium Resolution' (MPIM-ESM-MR) RCP 8.5 2050 values for mean sea temperature, salinity, primary production, molecular dissolved oxygen and ice cover for predicting future scenarios. Rather, debiasing is performed by taking the magnitude of change between the modelled future and modelled current conditions (e.g., SST2050), and applying this to present-day (current) conditions to get the environmental values (e.g., SST2050c) for the 2050 scenario. Note that the modelled future value (modelled 2050) of an environmental parameter is based on annual mean average for the years 2046-2055, whereas modelled current value (modelled 2007) is based on annual mean average for the years 2000-2014.

That is:

$$2050 \text{ debiased} = \text{Current} + (\text{Modelled 2050} - \text{Modelled 2007})$$

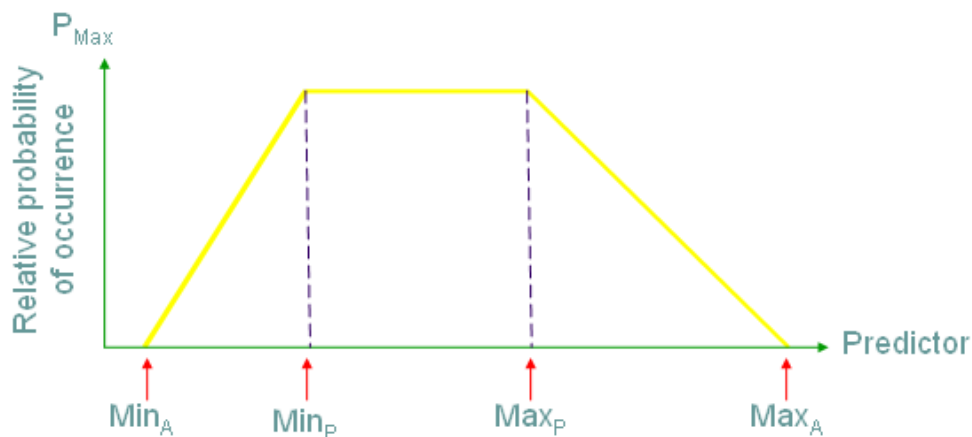
e.g.,

$$\text{SST2050c} = \text{SSTAnMean} + \text{SST2050}$$

Because this approach takes predicted change from models and superimposes it on real world data, it must be emphasized that this is just a rough hypothesis of future climate.

#### Environmental Envelopes

An environmental envelope is essentially a response curve that describes the habitat usage of a species or its preferences with respect to certain environmental factors. In AquaMaps, each species response curve has a trapezoidal shape (see Fig. 1). It assumes that the probability that a species is present is uniformly highest ( $P=1.00$ ) where mean environmental conditions fall within the preferred parameter range of the species ( $\text{Min}_P$  to  $\text{Max}_P$ ). If environmental conditions fall outside this range, the probability of occurrence is assumed to decrease linearly towards the species' absolute minimum or maximum parameter thresholds ( $\text{Min}_A$  or  $\text{Max}_A$ ), and is set to zero ( $P=0.00$ ) beyond the absolute threshold values.



**Figure 1. Model of a species-specific environmental envelope. A species will have an envelope for each of the environmental factors used to predict species occurrence.**

Absolute and preferred minima and maxima for all environmental factors (predictors), except for depth and sea ice concentration, are computed from the environmental attributes of “good cells” using the following rules:

1.  $Min_A = 25\text{th percentile} - 1.5 \times \text{interquartile}$  or absolute minimum in extracted data (whichever is lesser)
2.  $Max_A = 75\text{th percentile} + 1.5 \times \text{interquartile}$  or absolute maximum in extracted data (whichever is greater)
3.  $Min_P = 10\text{th percentile}$  of observed variation in an environmental predictor
4.  $Max_P = 90\text{th percentile}$  of observed variation in an environmental predictor

Sea ice concentration envelope threshold values are computed from the environmental attributes of “good cells” using the following rules:

1.  $Min_A = \text{absolute minimum in extracted data}$ . For all species where  $Min_A = 0$ , set  $Min_A$  to sea ice concentration mean value + (-1). This extends  $Min_A$  to avoid exclusion of species from all non-ice covered areas.
2.  $Max_A = \text{absolute maximum in extracted data}$
3.  $Min_P = 10\text{th percentile}$  of observed variation in an environmental predictor
4.  $Max_P = 90\text{th percentile}$  of observed variation in an environmental predictor

Depth envelope threshold values are taken directly from published data recorded in FishBase and SeaLifeBase.

Additional rules are also applied to prevent the use of nonsensical values or to conform to basic biological concepts. For instance:

- If  $\text{Depth } Min_A \leq 200 \text{ m}$ , envelope computations for sea temperature and salinity are based on surface values, and conversely, on bottom values if  $\text{Depth } Min_A > 200 \text{ m}$ .
- For all species with  $\text{Temperature } Max_P \geq 25 \text{ }^\circ\text{C}$ ,  $\text{Temperature } Max_A$  is set to  $\text{Temperature } Max_P + 4.2 \text{ }^\circ\text{C}$ . This sets the upper limit for tropical species to around  $34 \text{ }^\circ\text{C}$ , which is known to be the lethal limit for marine aquarium fishes.
- Where  $\text{Temperature } Max_A \leq 5 \text{ }^\circ\text{C}$  degrees (i.e., polar and deepwater species), the minimum distance between  $\text{Temperature } Min_P$  and  $\text{Temperature } Max_P$  should be  $0.25 \text{ }^\circ\text{C}$ , and  $1 \text{ }^\circ\text{C}$  where  $\text{Temperature } Max_A > 5 \text{ }^\circ\text{C}$ .
- Where  $\text{Primary Production } Min_P - Min_A < 1$ , set  $Min_A = Min_P - 1$ . This extends  $Min_A$  to avoid exclusion of species around the gyres.

Environmental envelopes for each species, along with the associated species model input data, are stored in a corresponding Species Environmental Envelope File (HSPEN), where species are assigned identification codes according to the Catalogue of Life: 2018 Annual Checklist

([catalogueoflife.org/annual-checklist/2018/](http://catalogueoflife.org/annual-checklist/2018/)). Species not yet listed here are assigned identification codes based on the corresponding synonym code of the currently accepted name of the species in FishBase (fishes) or SeaLifeBase (non-fishes).

#### Probabilities of Occurrence

The AquaMaps model computes for species-specific relative probabilities of occurrence by scoring how well environmental attributes (i.e., local conditions) in each half-degree cell match a species' environmental envelope.

The relative probabilities of occurrence range between 0.00-1.00. A probability of occurrence is first calculated for each of the individual environmental predictors (Fig. 1). Then the overall probability of species occurrence in a given half-degree cell ( $P_c$ ) is computed by taking the product of these individual probabilities.

There are however a few of exceptions. By default, AquaMaps computes  $P_c$  using all predictors except dissolved molecular oxygen and distance to land. Dissolved molecular oxygen is only included when computing  $P_c$  of deepwater species (i.e., DepthMin > 200 m). Distance to land, on the other hand, is optional and may be included as a restrictive buffer for species with specific life history traits that limit their occurrence away from the coast. Currently, it is only used to generate AquaMaps for several marine mammals, particularly some pinniped species that are known to be central place foragers. Also, in the case of marine mammals, mean depth in the cell (instead of cell depth range) is used in the computation of the probability of occurrence with respect to depth.

Thus, for species where Depth MinA  $\leq$  200 m:

$$P_c = P_{\text{depth}_c} \times P_{\text{surface temperature}_c} \times P_{\text{surface salinity}_c} \times P_{\text{primary production}_c} \times P_{\text{ice concentration}_c}$$

for species where Depth MinA > 200 m:

$$P_c = P_{\text{depth}_c} \times P_{\text{bottom temperature}_c} \times P_{\text{bottom salinity}_c} \times P_{\text{primary production}_c} \times P_{\text{ice concentration}_c} \times P_{\text{dissolved oxygen}_c}$$

and, inclusion of  $P_{\text{land distance}_c}$  is optional.

This multiplicative approach allows each environmental predictor to act as a “knock-out” criterion. For instance, if the sea temperature in a given cell exceeds the temperature preference of a particular species, the probability of occurrence with respect to temperature will be zero ( $P_{\text{temperature}_c} = 0.00$ ). The overall probability of occurrence of the species in that given cell will then be zero ( $P_c = 0.00$ ), even if all other environmental attributes of that cell are within the tolerance range of that species.

All non-zero overall probabilities of occurrence for a species and their associated half-degree cell (called CsquareCode) are stored in a corresponding Half-degree Species Assignment file (HSPEC). These records make up the map data which are sent to the C-squares Mapper ([www.cmar.csiro.au/csquares](http://www.cmar.csiro.au/csquares)) to generate the actual maps.

These computer-generated maps are marked as such in the title of the display page.



## Standard Map Outputs

The default set of AquaMaps for marine species consist of four basic maps (see Figs. 2a-2d):

**Native Range Map:** This map displays all areas with suitable environmental conditions that fall within a species' natural distributional range, as recorded in the literature (Fig. 2a). By default, the mapping algorithm ignores poleward boundaries imposed by FAO areas, because pole-ward extensions are better controlled by temperature preferences than by arbitrary management areas. In many cases, the predicted ranges will resemble a species' fundamental or historical niche rather than its realized or current native range, which may be impacted by other (anthropogenic) factors such as introductions, marine pollution, noise, and/or overfishing.

**Suitable Habitat Map:** This map disregards limits of the currently known geographic range of a species and displays all areas with suitable environmental conditions (Fig. 2b). If a species were introduced to an area outside its native range, this map shows the probability at which the species is likely to survive and establish itself there.

**Point Map:** The point map displays the available geo-referenced occurrence data (plotted in its corresponding half-degree cells) for a species, collected from GBIF, OBIS, FishBase and SeaLifeBase, as well as country points (Fig. 2c). Note that some of these points may be erroneous. A color legend is thus provided to identify "good" and "bad" cells, determined by filtering occurrence cells using a species' bounding box and/or FAO areas assignments. Cells based on country points are likewise differentiated by the legend.

**2050 map:** This map displays the possible range of natural distribution of a species by the year 2050 with respect to global climate change conditions described under the IPCC RCP 8.5 greenhouse gas concentration trajectory (Fig. 2d). The RCP 8.5 scenario describes a future world with high population and relatively slow income growth with modest rates of technological change and energy intensity improvement (Riahi et al., 2011). Because of these characteristics, it describes a scenario of increasing greenhouse gas emissions over time, leading to high greenhouse gas concentration levels peaking by year 2100.

The AquaMaps 2050 model assumes no changes in species tolerances and habitat usage. Hence, the calculation of relative probabilities of occurrence in 2050 uses the same species environmental envelopes as that for the current native range.

The predicted 2050 distribution allows either a 10° expansion in all four directions of the respective species bounding box, or an extension into the neighboring FAO areas in a pole-ward direction for cases where bounding box data are missing or incomplete. The expansion is meant to capture the shifts in a species' predicted range following changes in the marine environment under the RCP8.5 scenario. The 10° allowance for range expansion however is provisional and will be replaced when a more precise estimate has been determined.

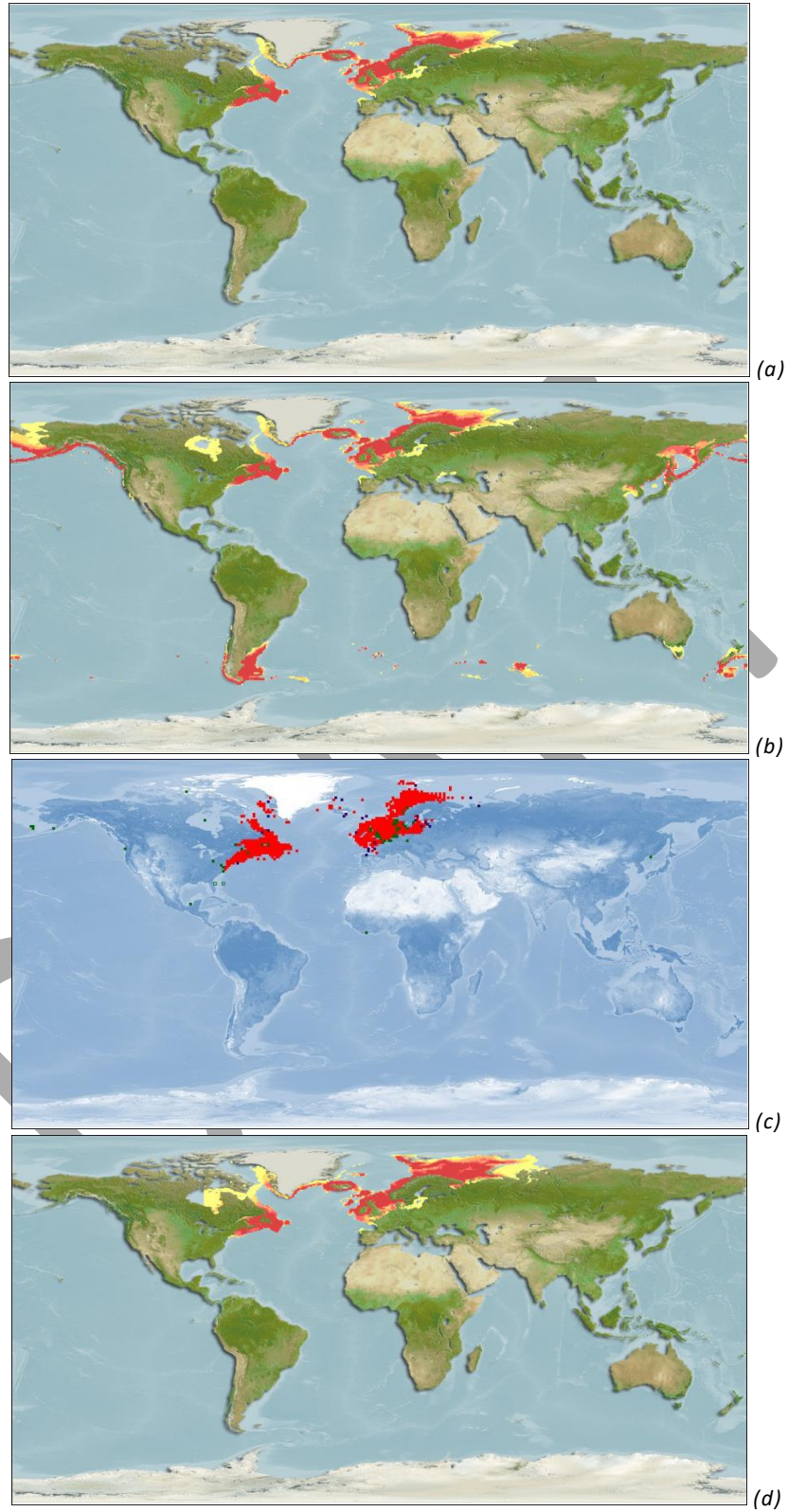


Figure 2. Standard marine AquaMaps for Atlantic Cod *Gadus morhua* Linnaeus, 1758; (a) native range map, (b) all suitable habitat map, (c) point map, and (d) 2050 native range map.

## *User-defined Maps*

The Create-Your-Own-Map (CYOM) interface provides an interactive mode where individual species maps can be edited by users. Through this interface, mapping parameters of a species can be edited, and maps can be regenerated to incorporate these changes. Mapping parameters can be modified in four ways:

### *1. Editing area restrictions of a species*

The selection of “good cells” and the computation of species environmental envelopes are functions of the area restrictions defined for a species. Corrections can be made to the following parameters:

- **FAO areas:** Add/delete FAO areas according to where a species is known to be native or endemic.
- **Pelagic flag:** Change setting to either TRUE if species distribution IS NOT influenced by bottom depth, or FALSE if species IS influenced by bottom depth.
- **Sea temperature and salinity layers:** Change to use either surface or bottom values.
- **Bounding box:** Adjust/complete latitudinal or longitudinal extents to encompass area of known native range of species.

“Good cells” and species’ environmental envelopes need to be re-calculated following any changes to the above parameters.

### *2. Adding/excluding cells used for creating environmental envelope*

The list of half-degree cells identified from occurrence data for a species includes both the “good cells” used in computing the default environmental envelopes for the species, and unused cells that fall outside the species’ bounding box or FAO area limits. Add “good cells” by manually entering geographic coordinates of occurrence points, or by selecting unused cells as appropriate. Exclude a “good cell” by deselecting the corresponding record from the list. Added or excluded cell records will be included/excluded upon envelope recalculation and will result in adjustments to the different thresholds of one or several environmental envelopes.

### *3. Adjusting minima and maxima of environmental envelopes*

Individual environmental envelopes may also be adjusted by manually changing minimum and maximum threshold values of the preferred and absolute ranges of a species. These changes must be saved before regenerating the map data.

### *4. Disabling/enabling environmental envelopes*

Individual environmental envelopes may also be excluded or included in the computation of relative probabilities of occurrence. This is done by selecting or unselecting the checkbox corresponding to each environmental envelope. Changes in the checkbox settings should likewise be saved before regenerating the map data.

Further information on how to use the tool can be obtained from the CYOM user manual that can be downloaded from the CYOM interface.

### *Reviewed Maps*

Computer-generated AquaMaps are reviewed to correct errors and improve default predictions. The AquaMaps team performs routine verification of maps by checking the predicted native range map of a species against its known distribution as stored in the FishBase and SeaLifeBase databases. These maps are also reviewed by species experts, typically by incorporating their knowledge about species occurrences and environmental tolerances. They may also evaluate predictions against independent data. All these are done using the CYOM interface described above. Maps that have undergone routine verification are tagged as “Team reviewed” while those inspected or corrected by an expert are tagged as “Expert reviewed”.

Species experts are required to register for free to save and display their improved maps in AquaMaps. Reviewed maps are accessed along with the latest computer-generated AquaMaps for a species. If a reviewed map for a species exists, it is displayed as the default species distribution map in the FishBase and SeaLifeBase Species Summary pages, otherwise the computer-generated map is displayed.

Expert-reviewed and routinely verified maps are flagged as “Reviewed” in the title of the map display page.

An expert who reviews a map and/or makes edits to the mapping parameters of a species needs to document the changes and their reasons and rate the quality of the map. This can be done by providing some brief notes in the Remarks field. The recommended content includes:

- Problem(s) with the previous version of the map.
- Action(s) taken/edits made to improve the map.
- References used as basis for corrections made, if any.
- Other important comments or notes.
- 

AquaMaps can also be rated using a 5-star rating scheme (see Table 3). A reviewed/edited map will not necessarily correspond to all conditions under each criterion within a given star rating and will most likely vary across star ratings with respect to the different criteria. Thus, these criteria for rating are best used as a guide to approximate the degree of reliability of the predicted species distribution in the expert-reviewed/edited map. The final rating is thus left to the discretion of the expert.

**Table 3. Five-Star rating scheme to guide an expert-reviewer in evaluating the reliability of an AquaMaps native range prediction for a given species.**

Criteria	Star Rating				
	5	4	3	2	1
<b>Environmental envelopes</b>	Envelopes ok; no further edits recommended	Envelopes ok; no further edits recommended	Envelopes ok although species known to have large inter-annual changes in habitat usage are only inadequately captured by single annual envelope	Envelopes ok but may still be improved by adjusting parameters (>2); or available environmental parameters are unable to adequately describe species occurrence	Computer-generated map
<b>Area restrictions</b>	Bounding box/basins complete and with good fit to known distribution	Bounding box/basins complete and with good fit to known distribution	Uses bounding box/basins; no further improvements of bounding box/basins possible but areas of false predicted presence remain	No bounding box/basins; defined by FAO areas that encompass entire known range	Computer-generated map
<b>Point data/good cells</b>	Adequately large sample size; samples cover representative portion of species range; no apparent bias introduced; no good cells needed to be added/removed	Adequately large sample size; samples cover representative portion of species range; biases were corrected by adding or removing good cells	Medium sample size and coverage of known species range but strong effort biases due to heterogeneous sampling effort; possible point data bias/output can only be improved with addition/deletion of a large number of good cells	Low sample size, non-representative coverage of species range by sampling	Computer-generated map
<b>Predicted range of occurrence</b>	In very good agreement with known range/significant statistical relationship between predictions and independent survey data	In good agreement with known range	Approximates known range but possibly with some areas of false predicted presence or absence	Approximates known range but includes large areas of false predicted presence or absence	Computer-generated map
<b>Predicted relative probabilities of occurrence</b>	In very good agreement with known relative occurrences /significant statistical relationship between predictions and independent survey data	In good agreement with known relative occurrences	Good correspondence with overall range but large discrepancies between predictions and known concentrations of high species occurrence	Good correspondence with overall range but large discrepancies between predictions and known concentrations of high species occurrence	Computer-generated map

## MARINE AQUAMAPS SEARCH PAGE

The Marine AquaMaps Search Page displays global *Marine Biodiversity Maps*. Radio buttons on top of the map allow the switching to several pre-defined groupings, currently for all species, sharks and rays, bony fishes, invertebrates, deep-sea species marine mammals, and open ocean fishes. Select a species group and click on any part of the oceans on the map to obtain a local list of species based on a probability of occurrence threshold of  $P > 0.5$  (used to represent a species' core habitat). Radio buttons can be used to toggle between lists showing native, potential invasives, or both. Some lists can be filtered further to show some pre-defined grouping, including: game fish, dangerous species, shallow water species, marine mammals, invertebrates, and algae.

Switching to the *List for Advanced Users* gives the user the option to customize the species list by specifying a different probability threshold. Additionally, a user can use the Biodiversity changes options to switch to the list of species predicted to be retained, lost or a new entrant in the area by the year 2050.

The header section of the Search Page gives a link to the latest *Environmental Data* which are available in the AquaMaps Half-degree Cell Authority File (HCAF/v.7, 10/2019). These are visualized as global maps of ocean depth (mean, minimum and maximum), as well as of current/observed and modelled 2050 data for mean surface and bottom sea temperature and salinity, mean primary productivity, mean sea ice concentration, and mean bottom dissolved oxygen. These are useful to understand the limits of species with regard to the environmental predictors used by the AquaMaps species distribution model.

Below the Marine Biodiversity Map, the Search Page offers several pre-defined search options: you can search for species maps by common or scientific name, view species richness maps for different taxonomic groups, or view checklists of species by Large Marine Ecosystem (LMEs) or by Country/Island. A variety of interactive Search Tools is also provided. Some of these options are explained below.

## AQUAMAPS TOOLS AND APPLICATIONS

**Advanced Search:** This search tool enables users to generate a checklist of species according to search category:

- by phylogeny (e.g., Kingdom, Phylum, Class, Order, Family)
- by special group (e.g., sharks, deep-sea fishes, fishes of Southeast Asia)
- by Large Marine Ecosystem (e.g., Baltic Sea, Kuroshio Current, Red Sea)

Each species on the list is linked to its corresponding set of standard AquaMaps as well as its reviewed maps, if available.

**Biodiversity Maps:** This tool allows users to view and/or generate multi-species maps by phylogeny, or large marine ecosystem. This search is applied to datasets where the probability of species occurrence is  $P > 0.5$ . The maps display the total number of species predicted to occur in a given half-degree cell. Species richness data are classified using a logarithmic scale displayed on the map in gradient of 10 colors from red (highest counts) to yellow (lowest counts).

**Climate Change Maps:** This is an advanced search routine that allows a user to view maps of predicted future range for a list of species by phylogeny or large marine ecosystem, alongside its corresponding current native range map. Other useful options include viewing the corresponding ocean area covered at the present and by the future range at a probability of occurrence of  $P > 0.5$ .

**Introduced Species Checklist:** This provides a list of marine fish species that have been introduced to one or more countries, with an indication of its invasiveness, and links to its corresponding species summary page in FishBase and maps in AquaMaps. A species' All Suitable Habitat map extends the predicted distribution beyond its native range to other areas in the ocean where conditions are suitable for it to survive and thus shows non-native habitats where an invasive species could potentially establish itself or that coincides with areas where a species has been introduced.

**Latitudinal (North-South)/Longitudinal (West-East) Species Richness Transects:** This tool allows a user to query species counts along a transect of half-degree cells based on a user-defined probability of occurrence (default  $P > 0$ ). Latitudinal transect are defined by a user by specifying a center longitude ( $\pm 0.25^\circ$  increments from  $0^\circ$ ), a starting latitude and an ending latitude. Longitudinal transect are defined by specifying a center latitude ( $\pm 0.25^\circ$  increments from  $0^\circ$ ), a starting longitude and an ending longitude. The query returns a list of half-degree cells by center latitude/longitude along with the species count, minimum depth (m) and ocean area ( $\text{km}^2$ ) corresponding in each half-degree cell. The list is accompanied by a line graph that plots the number of species along the latitudinal/longitudinal transect.

An equal-area version of the latitudinal species richness transect is planned for implementation. Query results will return a species count per  $\text{km}^2$  by  $0.25^\circ$  latitudinal increments for six 222-km wide transects running along the contour of the east and west coasts of the Atlantic, Pacific and Indian Ocean.

**MPA Planning Tool:** This tool helps a user to select suitable sites for designating a marine protected area. It employs a wizard where the user is asked to select the large marine ecosystem (LME), exclusive economic zone (EEZ) or FAO area that the area of interest is a part of, then to select species from a list to be included in the proposed MPA with regards to:

- **status of threat:** according to IUCN classification (e.g., critically endangered, endangered, vulnerable, near threatened)
- **dependence on the ecosystem:** expressed as percentage of overlap of the species' native range with the area
- **resilience:** ability of a species to recover from low population densities, such as may be caused by overfishing or natural catastrophes (e.g., very low, low, medium, high)
- **fishery status:** less than 50% ( $\sim$  overfished) or less than 10% ( $\sim$  collapsed) of the highest landings reported for the species in the covered FAO area
- **other importance:** species of special interest (e.g., according to popularity)

Once the initial list of species for protection has been generated, it can be modified by adding or removing species from the list. From the species selected, a map is generated and color-coded based on the number of species predicted to occur in every half-degree cell of the selected area. The map is interactive. Clicking on a point on the map will provide the list of the selected species that are in that area, along with their respective probability of occurrence.

Users are reminded however that species listed under this tool are limited to those with AquaMaps and may therefore be incomplete in terms of species occurring in an area. Further, many of the underlying maps have not yet been verified by experts. Outputs of this tool should be checked against independent sources before being used in MPA planning activities.

**Reviewed Maps:** This tool lists all AquaMaps that have been checked for agreement with a species known native or endemic distribution and edited to correct map errors. The list contains the scientific name of the species, the reviewer, the review date, the reviewed/edited map, reviewer remarks, the number of valid point data used to generate the species' environmental envelopes, and a 5-star rating indicating the reviewer's assessment of the reliability of the map.

Criteria used for the star rating are provided in Table 3. Reviewed maps can either be Team reviewed or Expert reviewed. Team reviewed maps are checked and/or edited by the AquaMaps map validation team for consistency with distribution data in FishBase, SeaLifeBase and other species databases, or with published species distribution maps. They adequately capture the current knowledge about a species' overall geographic range and at least some of the environmental envelope thresholds may have been verified using independent published information. The adequacy of the sample size of the point data used, and the agreement of predicted probability of occurrence with known relative occurrences of the species are generally not considered in the map rating given.

Expert-reviewed maps have been examined and/or edited by a species' expert or specialist of the species group. The review of the expert typically considers the correctness of the environmental envelopes in describing species tolerances, the adequacy of the sample size and representativeness of point data used, and the agreement of the predicted maximum range extent and relative probability of occurrence within that range with the species' known presence and relative occurrences.

**Predicted Change in Habitat Suitability (IUCN Threatened Species):** This tool illustrates how the modelled spatial distributions of IUCN Red listed marine species of threat status vulnerable (VU), endangered (EN), or critically endangered (CR) may change, under the IPCC RCP 8.5 scenario. Clickable thumbnails provide access to the modelled distribution of a species' currently known native range, its predicted range in year 2050, and a derived map showing areas in the species' distributional range that will retain the same suitability, have increased, or decreased in suitability, are no longer suitable, or are new suitable areas for the species. Area estimates of these predicted changes are provided in square kilometers and in corresponding half-degree cell counts. This enables users to easily see the most to the least predominant changes in habitat suitability within the distributional range of the species.

## DOWNLOADING DATA

AquaMaps is free and can be downloaded from [aquamaps.org](http://aquamaps.org). The available data sets and corresponding file formats are listed in Table 4 below.



**Table 4. Downloadable Marine AquaMaps Datasets (v.10/2019)**

Dataset	Comma-separated value file (csv)	netCDF (nc)	Excel file (xls)
Species range maps			
Probabilities of species occurrence (HSPEC)	Current native range	Current native range (computer-generated)	
Species environmental envelopes (HSPEN)	2050 range		
Occurrence cells	All Suitable Habitat (computer-generated, reviewed)		
Half-degree Cell Authority File (HCAF)			
Environmental data	x		
Metadata			x
Biodiversity Maps			
Species Richness	x		
Checklists			
Large Marine Ecosystems (LME)	x		
Country/Island	x		

## OTHER USER RESOURCES

The Hints link found at the upper section of the Species Search Page provides users with tips on how to make better use of AquaMaps. Topics covered include:

- Web browser settings
- Downloading and using maps
- Deep-linking to AquaMaps biodiversity species list and species maps

The Terms & Conditions link also found at the upper section of the Species Search Page provides additional information with respect to:

- Citing AquaMaps
- Using full or large set of AquaMaps data
- Privacy policy
- Disclaimer
- Copyright information

## MARINE AQUAMAPS STATISTICS

In the latest marine AquaMaps version (v. 10/2019), we have:

33518	total maps for marine species
12939	marine fishes
123	marine mammals
20056	other marine metazoans (=Kingdom Animalia and not Fish and not Class Mammalia)
299	macroalgae and marine vascular plants
66	biodiversity maps by pre-defined phylogenetic groups
66	checklists by LMEs

## INTERNET

Marine AquaMaps is available on the Internet and can be accessed directly through [www.aquamaps.org](http://www.aquamaps.org). It can also be accessed through the AquaMaps link or by clicking on the map thumbnail in the FishBase or SeaLifeBase species summary pages.

## REFERENCES

- Anderson, R.P., Lew, D. & Peterson, A.T. (2003). Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Modell.* 162, 211–232, doi:10.1016/S0304-3800(02)00349-6.
- Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2017). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*. 27(3), 277-284. doi: <https://doi.org/10.1111/geb.12693>.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H. D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., . . . Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3), 572–597. <https://doi.org/10.1002/jame.20038>
- Hastie, T.J. (1991). Generalized additive models. In: J.M. Chambers & T.J. Hastie (Eds.), *Statistical Models in S* (pp. 249-308). Wadsworth & Brooks/Cole.
- Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). The global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, n/a. <https://doi.org/10.1002/jame.20017>
- Kaschner, K., Watson, R., Trites, A., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, 316, 285–310. <https://doi.org/10.3354/meps316285>
- Kesner-Reyes, K., Segschneider, J., Garilao, C., Schneider, B., Rius-Barile, J., Kaschner, K., & Froese, R. (2019, October 14). AquaMaps Environmental Dataset: Half-Degree Cells Authority File (HCAF ver.7, 10/2019). Retrieved from [https://www.aquamaps.org/main/envt\\_data.php](https://www.aquamaps.org/main/envt_data.php)

- McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall.
- Pauly D., Zeller D., & Palomares, M.L.D. (Eds.), (2020). *Sea Around Us Concepts, Design and Data*. Retrieved from [www.seaaroundus.org](http://www.seaaroundus.org)
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Ready, J., Kaschner, K., South, A. B., Eastwood, P. D., Rees, T., Rius, J., Agbayani, E., Kullander, S., & Froese, R. (2010). Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, 221(3), 467–478. <https://doi.org/10.1016/j.ecolmodel.2009.10.025>
- Rees, T. (2002, May). "C-squares" - a new metadata element for improved spatial querying and representation of spatial dataset coverage in metadata records. EOGEO 2002 Proceedings, Ispra, Italy. <http://eogeo.jrc.it/>
- Rees, T. (2003). "C-Squares," A New Spatial Indexing System and its Applicability to the Description of Oceanographic Datasets. *Oceanography*, 16(1), 11–19. <https://doi.org/10.5670/oceanog.2003.52>
- Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., & Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, 109(1–2), 33–57. <https://doi.org/10.1007/s10584-011-0149-y>
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2011). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, 21(2), 272–281. <https://doi.org/10.1111/j.1466-8238.2011.00656.x>

**Kathleen Kesner-Reyes, Cristina Garilao, Kristin Kaschner, Josephine Barile, and Rainer Froese**

Cite as:

Kesner-Reyes, K., Garilao, C., Kaschner, K., Barile, J., & Froese, R. (2020). AquaMaps: algorithm and data sources for marine organisms. In: R. Froese & D. Pauly. (Eds.), *FishBase*. <https://www.fishbase.org>, version (10/2019).