

# *AquaMaps: Algorithm and Data Sources for Aquatic Organisms*

## AquaMaps for Marine Species

AquaMaps is a tool for generating model-based, large-scale predictions of natural occurrences of species. For marine species, the model uses estimates of environmental preferences with respect to depth, water temperature, salinity, primary productivity, 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 ([www.gbif.org](http://www.gbif.org)) and OBIS ([www.iobis.org](http://www.iobis.org)), and from independent knowledge from the literature about the distribution of a given species and its habitat usage that are available in FishBase (and in SeaLifeBase and AlgaeBase for non-fish). 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 by Kaschner et al. (2006) for predicting global distributions of marine mammals. 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. A description of the use of AquaMaps for freshwater organisms is given below.

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 - generalised 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 expert-reviewed maps can, from then on, only be updated by experts. In contrast, the computer-generated maps are updated every 1-2 years, as new data become available.

## ***Computer-generated AquaMaps***

### ***Input Parameters***

The process of generating the map data 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 Biogeographic Information System), as well as from museum collections and the literature available in FishBase and SeaLifeBase. Since 2010, we also use shallow and deep water “country points”, i.e., coordinates assigned to a country, which are likely to be inhabited by species reported to occur in that country. These point data are used to identify half-degree cells in an authority file from which environmental parameter values will be extracted to generate the species envelope.

**Bounding Box:** This consists of manually encoded geographic coordinates that delimit the known 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 on the literature. The bounding box is used to identify “good cells”, i.e., those half-degree cells which contain occurrence data and which are located within the bounding box. This is done to prevent processing of “bad cells” resulting from errors in the 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, although 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 have been assigned to these areas. For the purpose of 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 needed 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 indicative of the absolute (min–max) and the common depth range 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, “most likely” 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 applied to a given species.

**Pelagic Flag:** This flag indicates whether the temperature and salinity preferences of a species refer to surface or bottom values. Entries in the SPECIES table regarding the environment in relation to 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 SPECIES table.**

Environment/Province	Pelagic Flag
Pelagic-neritic	False
Pelagic-oceanic	True
Demersal	False
Benthopelagic	False
Reef-associated	False
Bathypelagic	True
Bathyemersal	False

### ***Environmental Parameters***

The AquaMaps algorithm for marine species uses six environmental parameters representing key physical and biological factors that structure the habitat of many species at large scales. Geo-referenced values for these parameters 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://www.loicz.org>) and C-squares ID numbers (<http://www.marine.csiro.au/csquares/>). The HCAF contains such environmental attributes for a grid of 179,904 half-degree cells over marine waters. Each cell contains environmental parameter values for:

**Depth:** Refers to minimum and maximum cell bathymetry derived from ETOPO 2min negative bathymetry elevation; in meters.

#### **Temperature:**

- Observed mean annual surface and bottom sea temperature derived from NCEP SST Climatology (1982-1999); in degrees Celsius.
- Modeled current mean annual surface and bottom sea temperature from the IPSL Climate model SRES A2 (2001-2010); in degrees Celsius.
- Modeled 2100 mean annual surface and bottom sea temperature from the IPSL Climate model SRES A2 (2090-2099); in degrees Celsius.

#### **Salinity:**

- Observed mean annual surface salinity provided by the World Ocean Atlas (1982-1999); in PSU.
- Observed mean annual bottom salinity provided by the World Ocean Atlas Bottom Source Information (1990-1999); in PSU.
- Modeled current mean annual surface and bottom salinity from the IPSL Climate model SRES A2 (2001-2010); in PSU.
- Modeled 2100 mean annual surface and bottom salinity from the IPSL Climate model SRES A2 (2090-2099); in PSU.

**Primary Production:**

- Proportion of annual primary production in a cell from [http://searounds.org/PrimaryProduction/Interpolation\\_method.htm](http://searounds.org/PrimaryProduction/Interpolation_method.htm); in  $\text{mgC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ .
- Modeled proportion of annual primary production in a cell from the IPSL Climate model SRES A2 (2001-2010); in  $\text{mgC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ .
- Modeled 2100 proportion of annual primary production in a cell from the IPSL Climate model SRES A2 (2090-2099); in  $\text{mgC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ .

**Sea Ice Concentration:**

- Observed mean annual ice cover as derived from the National Snow and Ice Data Centre (1979-2002), <http://nsidc.org/data/nsidc-0051.html>; in percent.
- Modeled current mean annual ice cover from the IPSL Climate model SRES A2 (2001-2010); in percent.
- Modeled 2100 mean annual ice cover from the IPSL Climate model SRES A2 (2090-2099); in percent.

The AquaMaps model does not make straightforward use of modeled 2100 values for mean annual sea temperature, salinity, primary production and ice cover for predicting future scenarios. When looking at the difference between observed and modeled environmental datasets (current and 2100), a large part of the change in environment actually does not have to do with the degree of predicted change in an environmental parameter, but rather with the difference between the modeled environment and the real world. AquaMaps therefore takes the difference between both modeled environments and applies this to the existing observed environment to get a new environmental dataset for the future scenario. That is:

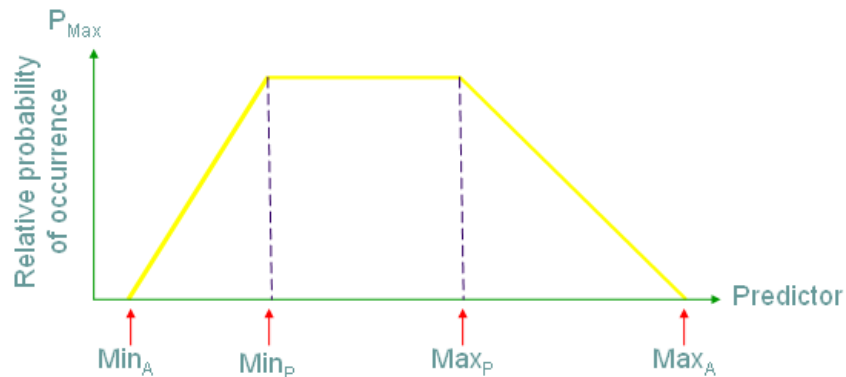
$$2100 \text{ adjusted} = \text{Observed} + (\text{Modeled } 2100 - \text{Modeled current})$$

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

**Distance to Land:** distance of marine cell to nearest coastal cell; in kilometers.

## ***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 parameters. 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 parameters used to predict its occurrence.**

Absolute and preferred minima and maxima for all environmental parameters, except for depth, are computed from the environmental attributes of “good cells” using the following rules:

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

Depth envelope threshold values are taken directly from data 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  $\text{Depth Min}_A \leq 200$  m, envelope computations for sea temperature and salinity are based on surface values, and conversely, on bottom values if  $\text{Depth Min}_A > 200$  m.
- For all species with  $\text{Temperature Max}_P \geq 25$  °C,  $\text{Temperature Max}_A$  is set to  $\text{Temperature Max}_P + 4.2$  °C. This sets the upper limit for tropical species to around 34 °C, which is known to be the lethal limit for marine aquarium fishes.

- Where Temperature MaxA  $\leq 5$  °C degrees (i.e. polar and deepwater species), the minimum distance between Temperature MinP and Temperature MaxP should be 0.25 °C, and 1 °C where Temperature MaxA  $> 5$ °C.

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

### ***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 a given half-degree cell match a species' environmental envelope.

The relative probabilities of occurrence are presented as probability values ranging between 0.00-1.00. A probability of occurrence is first calculated for each of the individual environmental predictors. The product of these individual probabilities is then taken to determine the overall probability of occurrence ( $P_c$ ) in a given half-degree cell, that is:

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

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 preferred range of that species.

By default, AquaMaps computes for the overall probability of occurrence in a cell using all parameters except distance to land. The use of this parameter 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, for marine mammals mean depth in the cell (instead of cell depth range) is used in the computation of the overall probabilities of occurrence.

All non-zero 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 (<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 known from the literature. By default, the mapping algorithm ignores pole-ward boundaries imposed by FAO areas, because pole-ward extensions are better controlled by temperature preferences than by arbitrary management areas. In many cases, the predictions will resemble a species' fundamental or historical niche rather than its realized or current native range, which may be impacted by other factors such as introductions 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. If the species were introduced to such an area outside its native range, this map shows the probability at which the species is likely to survive and establish itself there.

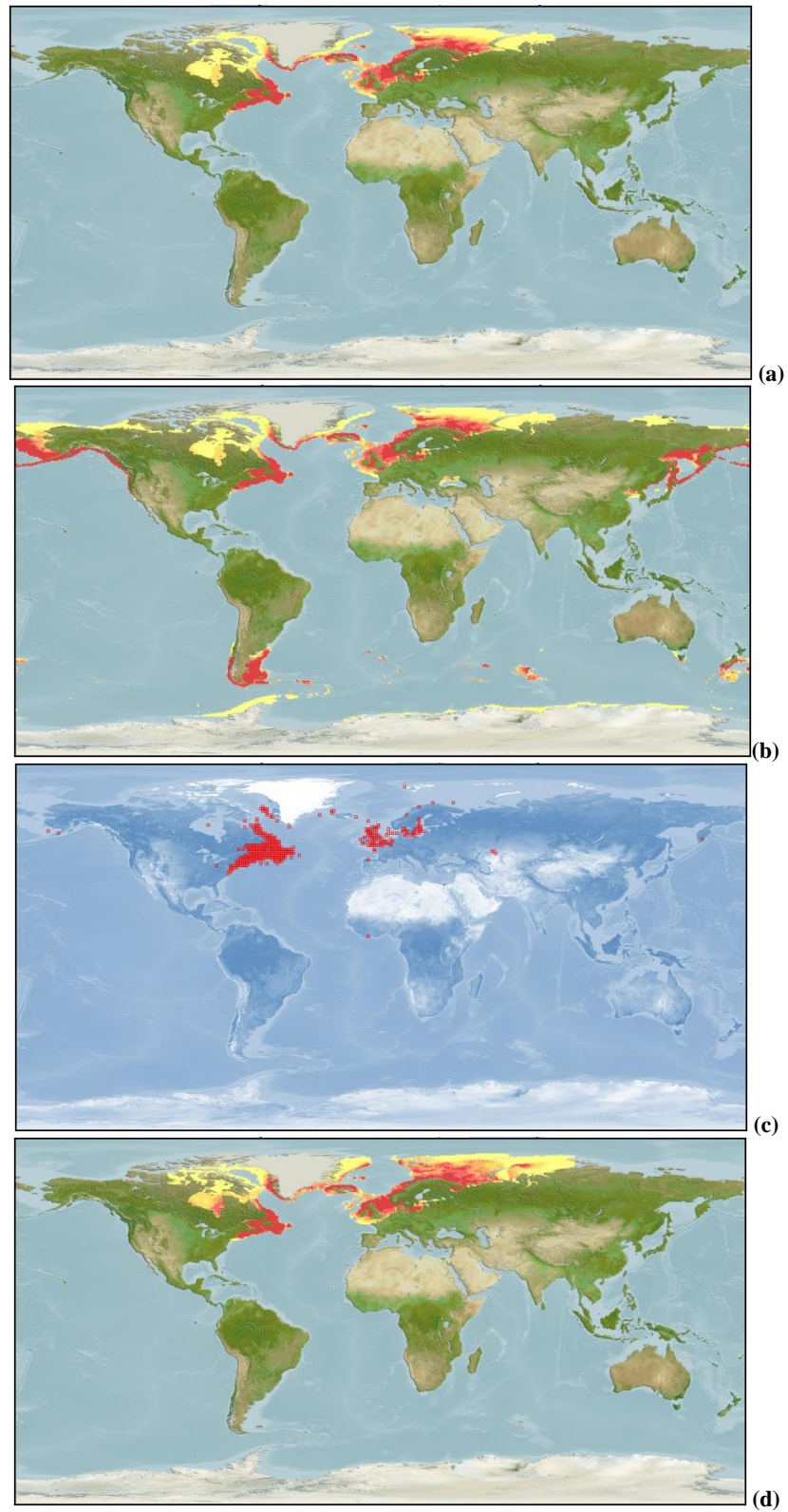
**Point Map:** The point map displays the available geo-referenced occurrence data (plotted in its corresponding half-degree cells) for the species, collected from GBIF, OBIS, FishBase and SeaLifeBase. Note that some of these points may be erroneous. A color legend is thus provided to identify “good”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

**2100 map:** This map displays the possible range of natural distribution of a species by the year 2100 with respect to global climate change conditions described under IPCC SRES A2 scenario. The A2 emission scenario describes a heterogeneous future world consisting of independently operating, self-reliant nations, with a continuously increasing population. Furthermore, economic development is said to be regionally-oriented, and technological change slower and more fragmented than other IPCC SRES storylines (IPCC, 2000).

The AquaMaps 2100 model assumes no changes in species tolerances and habitat usage. Hence, the calculation of relative probabilities of occurrence in 2100 uses the same species environmental envelopes as that for the current native range. As sea level rise is assumed to be negligible, the depth envelope is assumed to remain the same as at present.

The predicted distribution allows either a 10° expansion in all four directions of the respective species bounding box, or extends 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 A2 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 AquaMaps for Atlantic Cod *Gadus morhua* Linnaeus, 1758; (a) native range map, (b) suitable habitat map, (c) point map, and (d) 2100 native range map.**



## *User-defined Maps*

The Create Your Own Map (CYOM) interface provides an interactive mode where maps can be edited by users. Through this interface, mapping parameters for a species can be changed and default 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 native or endemic.
- **Pelagic flag:** Change setting to either TRUE if species distribution is not affected by depth, or FALSE if it 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.

## ***Reviewed Maps***

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

Experts are required to register for free with FishBase and SeaLifeBase in order to save and display their expert-reviewed maps in AquaMaps. Expert-reviewed maps are listed along with the latest computer-generated AquaMaps for a species. If a reviewed map exists for a species, 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 marked 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 2). 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 2. 5-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
Environmental envelopes	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
Area restrictions	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
Point data/good cells	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
Predicted range of occurrence	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
Predicted relative probabilities of occurrence	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

## AquaMaps Search Page

The AquaMaps Search Page displays global Species Richness 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 the local list of species.

The header section of the Search Page gives a link to the Environmental Data which are available in the AquaMaps Half-degree Cell Authority File (HCAF). The data are visualized in global maps. This is useful to understand the limits of species with regard to the used environmental parameters.

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

## AquaMaps Tools and Applications

**Species Checklists:** Pre-defined and advanced searches are available to enable users to generate a checklist of species according to several search categories:

- 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)
- by country or island/territory

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

**Biodiversity Maps:** A number of pre-defined Species Richness Maps are available from radio buttons on top of the Search Page. These pre-defined maps are interactive. Clicking on a point on a map will, by default, provide a list of species predicted in that area based on a default probability of occurrence threshold of  $P > 0.5$ . The list can also be changed through the Species radio buttons to toggle to the list of potential invasive species, or the list that includes both native and potential invasive species. These lists can be filtered further under Predefined grouping. Categories for this include: game fish, dangerous species, shallow water species, marine mammals, invertebrates, algae and all marine species.

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 2100.

**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 by the current and future native range at species probabilities of occurrence of  $P \geq 0.5$ .

**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 coincides with areas where a species has been introduced.

**Latitudinal/Longitudinal 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 (e.g.,  $P > 0.2$ ). 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 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 this 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% (~ overfished) or less than 10% (~ 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 suggested for protection has been generated, it can be modified by adding or removing species from the list. From the species selected, a map of the area is generated and color-coded based on the number of the selected species predicted to occur in a given locality. The map is interactive. Clicking on a point on a map will provide a 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. Thus, results of this tool are still best verified against independent sources before being used for MPA planning.

**Web Service:** This service allows the inclusion of AquaMaps into a user's own web page. A user can choose from two URLs which can be edited with respect the genus name and species name to get an XML output that can be copied and pasted into the user's web page. The interface includes a link to a wrapper program that can be downloaded for this service.

### **Marine AquaMaps Statistics**

As of August 2015 we have:

- 22889 total maps for marine species
- 12068 marine fishes
- 118 marine mammals
- 10159 other marine metazoans (=Kingdom Animalia and not Fish and not Class Mammalia)
- 116 biodiversity maps by pre-defined phylogenetic groups
- 66 checklists by LMEs
- 240 checklists by country or island/territory

### **Internet**

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

## **AquaMaps for Freshwater Species**

Freshwater AquaMaps is a tool for generating model-based, large-scale predictions of natural occurrences of species over continental basins. The model uses estimates of environmental preferences in terms of elevation, surface temperature, net primary productivity, soil pH, soil moisture, soil organic carbon, precipitation and compound topographic index. These estimates of species preferences, called environmental envelopes, are derived from large sets of occurrence data available from GBIF ([www.gbif.org](http://www.gbif.org)), and museum survey collections, and from independent knowledge from the literature about the distribution of a given species and its habitat usage that are available in FishBase. The environmental envelopes are matched against local environmental

conditions to determine the suitability of a given area for a particular species. Predictions of relative probabilities of species occurrence are shown as color-coded species range maps with a resolution 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 combining species occurrences and expert knowledge to derive the environmental preferences of a species is adapted by FishBase from a relative environmental suitability model originally developed by Kaschner et al. (2006), and was first implemented by FishBase for predicting global distributions of marine organisms, and subsequently freshwater fishes. The approach specifically corrects for biases in occurrence data which typically include non-representative coverage of a species' distribution, biases in sampling effort and data provision, and species misidentifications. A description of AquaMaps for marine organisms is given in the preceding section.

The AquaMaps tool for freshwater fishes was developed in collaboration with the Swedish Museum of Natural History (NRM). Although initially applied to predicting natural occurrences of fishes in the Americas, efforts to generate predictions of natural occurrences of freshwater fishes in Europe, Africa and China have also been made.

### **How Freshwater AquaMaps Works**

AquaMaps predictions of freshwater species distribution can be generated from two processes. In the first process, 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. In particular, the AquaMaps algorithm uses bounding basins and FAO areas occupied by the species. In a second process, computer-generated maps are reviewed, edited and/or approved by experts. Expert-reviewed maps can only be updated by experts and can be done at any time, while the computer-generated maps are updated every 1-2 years, as new data become available.

#### ***Computer-generated AquaMaps***

##### ***Input Parameters***

The process of generating the map data begins with the consolidation of the following input data:

**Occurrence data:** These consist of geo-referenced occurrence records harvested from the GBIF (Global Biodiversity Information Facility (GBIF) as well as from museum collections, and the literature available in FishBase. These point data are used to identify half-degree cells in an authority file from which environmental parameter values will be extracted to generate the species environmental envelopes.

**Bounding basins:** These are drainages where a species naturally occurs. They are determined from species-ecosystem-basin assignments, derived by linking (1) records in FishBase of fresh-



water ecosystems where a species is endemic or native, and (2) a reference table of freshwater ecosystems in FishBase and the corresponding Pfafstetter basins that encompass them. Pfafstetter basin levels used vary from Level 2 to Level 6, or can be a combination of these depending on the continent where the species distribution is being modelled. For instance, the AquaMaps systems for the Americas and China use Level 2 basins, whereas the system for Europe uses Level 3 basins (with a few hybrid basins). AquaMaps for Africa uses a combination of levels down to Level 6.

Bounding basins are used to identify “good cells” or half-degree cells within a bounding basin which contain occurrence data. This approach filters out errors in the occurrence data by excluding point data in half-degree cells that are outside of the known natural range of a species. The bounding basins are also used to prevent the display of the native range in areas of suitable habitat that are inaccessible to the species. For example, a species that is endemic to the African mainland will not be shown in Madagascar, despite having suitable habitat there.

Further information on the Pfafstetter drainage classification system and the Pfafstetter basins data sets are available from HYDRO1K, U.S. Geological Survey EROS Center at [http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data\\_Available/gtopo30/hydro](http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro).)

**FAO Areas:** These are the large geographical inland fishing areas designated by the Food and Agriculture Organization of the United Nations (FAO) for statistical purposes. All species in FishBase have been assigned to these areas. For the purpose of AquaMaps, only those FAO areas where a species is endemic or native are considered. Similar to the bounding basin, FAO areas are used to identify “good” cells and to exclude point data outside of the known range of a species. A minimum of 10 different “good” cells is needed to generate a map for a species.

### *Environmental Parameters*

The freshwater AquaMaps algorithm uses eight environmental parameters representing key physical and biological factors that structure the habitat of many freshwater species at large scales, and can be used as predictors of species occurrence. Geo-referenced values for these parameters at 0.5° resolution are stored as sets of cell attributes in a Half-degree Cell Authority File (HCAF) along with their associated LOICZ (<http://www.loicz.org>) and C-squares cell ID numbers (<http://www.marine.csiro.au/csquares/>). The HCAF contains such environmental attributes for at least 58,180 half-degree cells over land. Each cell contains environmental parameter values for:

**Elevation:** Elevation above seal level derived from ETOPO2, National Geophysical Data Center <http://www.ngdc.noaa.gov/mgg/global/relief/ETOPO2/ETOPO2v2-2006/>; in meters.

**Monthly temperature:** Average temperature of a grid cell during June derived from CRU 0.5 Degree Dataset, New et al.(1998), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=35&includerelatedlinks=1&dataset=35>; in

degrees Celsius. (Note: Monthly average temperature for June is currently used for Freshwater AquaMaps. Annual average temperature may also be used.)

**Net primary productivity:** Total net primary productivity (NPP) in each grid cell over one year derived from Foley et al. (1996), Kucharik et al. (2000), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=37&includerelatedlinks=1&dataset=37>; in  $\text{kg-C}\cdot\text{m}^{-2}\cdot\text{year}^{-1}$ .

**Soil pH:** Measure of the acidity or alkalinity (pH) of the soil in a grid cell derived from IGBP-DIS (1998), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=20&includerelatedlinks=1&dataset=20>; on logarithmic scale of 0 to 14. (Note: Soil pH is currently excluded by default in Freshwater AquaMaps for Europe.)

**Soil moisture:** Amount of water in the soil available to vegetation in a grid cell derived from Willmott and Matsuura (2001), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23>; in millimeters (based on 150 mm Water Holding Capacity).

**Soil organic carbon:** Total mass of soil carbon (to one meter) contained in a grid cell derived from IGBP-DIS (1998), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=21&includerelatedlinks=1&dataset=21>; in  $\text{kg-C}\cdot\text{m}^{-2}$  to 1m depth.

**Annual total precipitation:** Total amount of precipitation that falls in one grid cell during an average year derived from CRU 0.5 Degree Dataset, New et al.(1998), Atlas of the Biosphere <http://www.sage.wisc.edu/atlas/maps.php?datasetid=34&includerelatedlinks=1&dataset=34>; in millimeters per year.

**Compound Topographic Index (CTI):** Wetness Index; function of the upstream contributing area and the slope of the landscape derived from HYDRO1K, U.S. Geological Survey EROS Center [http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data\\_Available/gtopo30/hydro](http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro); on scale of 0 to over 10. (Note: CTI is currently excluded by default in Freshwater AquaMaps for Europe.)

### ***Environmental Envelopes***

An environmental envelope is a species response curve that describes the tolerance of a species with respect to a given environmental parameter. As with AquaMaps for marine species, an environmental envelope for a freshwater species is trapezoidal in shape (see Fig. 1, p.5). It assumes that the probability of species occurrence is equal to 1 ( $P=1.00$ ) where average local conditions in an area are within the preferred range of a species ( $\text{Min}_p$  to  $\text{Max}_p$ ) for a particular environmental parameter. The probability of occurrence is assumed to decrease linearly beyond the thresholds of preferred range towards the absolute minimum and maximum tolerance thresholds of a species

(Min<sub>A</sub> or Max<sub>A</sub>). Beyond the absolute tolerance thresholds, the probability of species occurrence is equal to zero (P=0.00).

Absolute and preferred minima and maxima for all environmental parameters are computed from the environmental attributes of “good cells” using the following rules:

1. MinA = 25th percentile - 1.5 × interquartile or absolute minimum in extracted data (whichever is lesser)
2. MaxA = 75th percentile + 1.5 × interquartile or absolute maximum in extracted data (whichever is greater)
3. MinP = 10th percentile of observed variation in an environmental parameter
4. MaxP = 90th percentile of observed variation in an environmental parameter

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

### ***Probabilities of Occurrence***

The AquaMaps model computes the relative probabilities of occurrence by comparing local conditions in a given half-degree cell (represented by environmental attributes of a given cell) against the environmental envelopes of a given species. A probability of occurrence is first calculated for each of the individual environmental parameters used to model a species' distribution. Probability values range between 0.00-1.00. Then, the product of the probabilities for each of the environmental parameters is taken to compute the overall probability of species occurrence (P<sub>c</sub>) in a given half-degree cell, that is:

$$P_c = P_{\text{elevation}_c} \times P_{\text{temperature}_c} \times P_{\text{NPP}_c} \times P_{\text{soil pH}_c} \times P_{\text{soil moisture}_c} \times P_{\text{soil carbon}_c} \times P_{\text{precipitation}_c} \times P_{\text{CTI}_c}$$

Through this multiplicative approach, each environmental parameter acts as a “knock-out” criterion. For instance, if the elevation in a given cell is outside of the range of elevation at which a given species occurs, the probability of occurrence with respect to elevation will be zero (P<sub>elevation<sub>c</sub></sub> = 0.00). The overall probability of occurrence of the species in that given cell will therefore be zero (P<sub>c</sub> = 0.00), even if all other environmental attributes of that cell are within the preferred range of the species.

All non-zero probabilities of occurrence and their associated half-degree cell CsquareCode are stored in a Half-degree Species Assignment file (HSPEC) for a given species. These records make up the map data which are sent by to the C-squares Mapper (<http://www.cmar.csiro.au/csquares>) to generate the visible AquaMaps. The computer-generated maps are marked as such in the title of the display page.

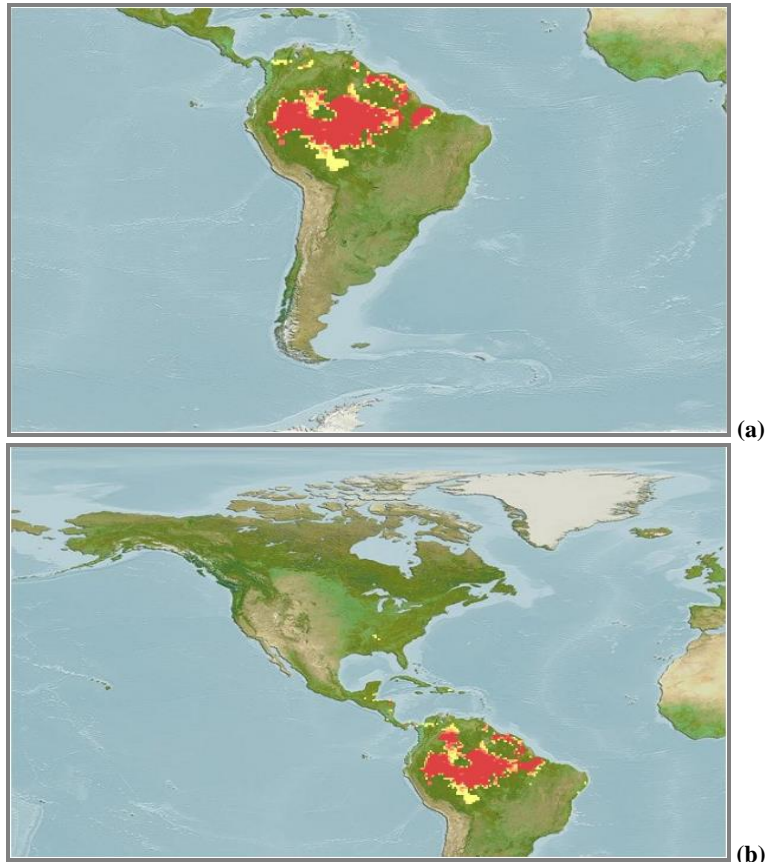
### *Standard Map Outputs*

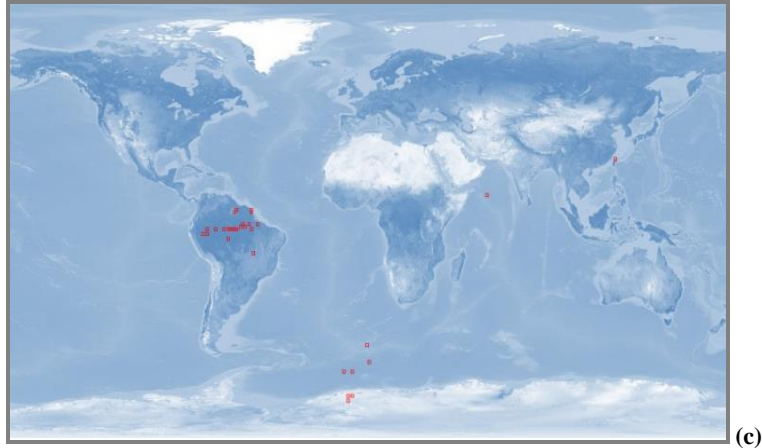
The default set of AquaMaps for freshwater species consist of three basic maps (see Figs.3a-3c):

**Native Range Map:** This map displays all areas with suitable environmental conditions that fall within a natural distributional range of a species, as known from the literature.

**Suitable Habitat Map:** This map disregards limits of currently known geographic range of a species. All areas with suitable environmental conditions aside from the native range of a species are also displayed. A species introduced to suitable areas outside its native range is likely to survive and establish itself.

**Point Map:** The point map displays available geo-referenced occurrence data (plotted in its corresponding half-degree cell) for the species, collected from GBIF, museum collections and FishBase. Note that some of these points may be erroneous. Future scenarios on the distributional range of freshwater species are not yet available in the current version of Freshwater AquaMaps.





**Figure 3. Standard AquaMaps for the arawana *Osteoglossum bicirrhosum* (Cuvier, 1829); (a) native range map, (b) suitable habitat map, and (c) point map.**

### *User-defined Maps*

The Create Your Own Map (CYOM) interface provides an interactive mode where maps can be edited by users. Through this interface, mapping parameters for a species can be changed and default 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 native or endemic.
- **Basins:** Add/delete Pfafstetter basins to approximate the area that encompasses the known native range of a species.

“Good cells” and the environmental envelopes need to be re-calculated following any changes made to the FAO areas and/or Basins assignments of a species.

#### *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 basins 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 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.

### ***Reviewed Maps***

Computer-generated AquaMaps are reviewed to correct errors and improve default predictions. Default maps can be review either through routine verifications performed by FishBase against species distribution information in the database, or through examinations made by species experts. Modification of the maps under review is likewise done through the CYOM interface. Reviewers are required to register for free with FishBase to be able to edit and display reviewed maps on the AquaMaps web interface. Reviewed maps are rated by the reviewer to give an indication of the reliability of the predicted species distribution shown in the reviewed/edited map. More details on reviewed maps are provided on pages10-11.

### **Freshwater AquaMaps Search Page**

While a single web-based Freshwater AquaMaps web interface similar to the marine interface is being developed, current access to freshwater AquaMaps is via four websites, corresponding to the predictive maps generated for the Americas, Europe, Africa and China. The upper section of each Search Page features a map of species richness predicted for the corresponding continent or that for China. The Species Richness Maps however are based on a limited and variable set of test species since the Freshwater AquaMaps system is still at an early stage of development relative to Marine AquaMaps. The latter is a more advanced system featuring 2100 predictions, and offering more search options, several tools/applications as well as web services.

The Species Richness Maps in the middle of search page are interactive. Click anywhere in the map where there is an estimate of species richness to get the list of species predicted in the half-degree cell selected. The list contains the number of species predicted to occur in the area with a probability of occurrence greater than 0.5, as well as their corresponding scientific name, common name, a picture of the species and their corresponding predicted native range map. The bottom of the list provides the latitude and longitude of the half-degree cell clicked, along with search fields that allow the user to search for an area by specifying a particular latitude and longitude in decimal degrees.

Below the Species Richness Map, the Search Page offers a search for species maps either by common or by scientific name.

The Tools section below the Scientific Name search fields features the Freshwater Checklist tool which generates a list of fishes reported in a given ecosystem, along with their corresponding native range AquaMaps. This tool however is only available for Freshwater AquaMaps for the Americas and only for selected ecosystems in the region.

### **Freshwater AquaMaps Statistics**

Fish species with AquaMaps as of 2015:

- 688 AquaMaps for the Americas (v. 07/2009)
- 258 AquaMaps for Africa (v. 02/2014)
- 33 AquaMaps for Europe (v. 09/2013)
- 32 AquaMaps for China (v. 09/2011)

### **Internet**

Freshwater AquaMaps is available on the Internet. Currently available versions can be accessed through the following:

AquaMaps for the Americas: [http://www.aquamaps.org/AM\\_Americas/search.php](http://www.aquamaps.org/AM_Americas/search.php)

AquaMaps for Africa: [http://www.aquamaps.org/AM\\_Africa/search.php](http://www.aquamaps.org/AM_Africa/search.php)

AquaMaps for Europe: [http://www.aquamaps.org/AM\\_Europe/search.php](http://www.aquamaps.org/AM_Europe/search.php)

AquaMaps for China: [http://www.aquamaps.org/AM\\_China/search.php](http://www.aquamaps.org/AM_China/search.php)

Alternative access is available through [www.aquamaps.org](http://www.aquamaps.org). The **Freshwater AquaMaps** link at the upper section of the Search Page launches the Freshwater AquaMaps Home Page where links to the search pages for the Americas, Africa, Europe and China are displayed.

### **References**

- Anderson, R.P., D. Lew, and A.T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Model.* 162, 211–232, doi:10.1016/S0304-3800(02)00349-6.
- Foley, J.A., I.C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine (1996) An Integrated Biosphere Model of Land Surface Processes, Terrestrial Carbon Balance and Vegetation Dynamics, *Global Biogeochemical Cycles*, 10, 603-628.
- Hastie, T.J., 1991. Generalized additive models. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical Models*. S, Wadsworth & Brooks/Cole, Chapter 7.
- IGBP-DIS (1998) *SoilData(V.0)* A program for creating global soil-property databases, IGBP Global Soils Data Task, France.



- International Panel on Climate Change (2000). IPCC Special Report Emissions Scenarios. Summary for Policymakers. 27 p.
- Kaschner, K., R. Watson, A.W. Trites and D. Pauly. 2006. Mapping worldwide distributions of marine mammals using a Relative Environmental Suitability (RES) model. *Mar. Ecol. Prog. Ser.* 316:285-310.
- Kucharik, C.J., J.A. Foley, C. Delire, V.A. Fisher, M.T. Coe, J. Lenters, C. Young-Molling, N. Ramankutty, J.M. Norman, and S.T. Gower (2000): Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance and vegetation structure. *Global Biogeochemical Cycles* 14(3), 795-825.
- McCullagh, P., and J.A. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall.
- New, M.G., M. Hulme and P.D. Jones, 1999: Representing 20th century space-time climate variability. I: Development of a 1961-1990 mean monthly terrestrial climatology. *J. Climate*. 12, 829-856.
- Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modelling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- 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. *Ecol. Model.* 221: 467-478, doi:10.1016/j.ecolmodel.2009.10.025
- Rees, T. 2002. "C-squares" - a new metadata element for improved spatial querying and representation of spatial dataset coverage in metadata records. EOGEO 2002 Proceedings, Ispra, Italy, May 2002. World Wide Web publication.
- Rees, T. 2003. "C-Squares", a New Spatial Indexing System and its Applicability to the Description of Oceanographic Datasets. *Oceanography* 16 (1), pp. 11-19.
- Willmott, Cort J., and Matsuura Kenji (2001): Terrestrial Water Budget Data Archive: Monthly Time Series (1950-1999).

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

Cite as:

- Kesner-Reyes, K., K. Kaschner, S. Kullander, C. Garilao, J. Barile, and R. Froese. 2016. AquaMaps: algorithm and data sources for aquatic organisms. In: Froese, R. and D. Pauly. Editors. 2012. FishBase. World Wide Web electronic publication. [www.fishbase.org](http://www.fishbase.org), version (04/2012).