\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**TITLE: AtlantECO-MAPS-v2-picozoa-annual-epipelagic\_habitat\_suitability\_index\_Huberetal2024\_20240507**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1.- INTRODUCTION**

The NetCDF files contain the extraplolated habitat suitability maps for 53 picozoa operative taxonomic units (pOTUs), described in the manuscript Huber et al. 2024 (submitted April 2024).

**2.- METHODOLOGY USED**

Habitat suitability maps were created for each pOTU using species distribution models (SDMs) with a presence/absence design. For each pOTU, all locations where a pOTU was detected were considered as “presences”, whereas locations where the pOTU was not detected were considered as “pseudo-absences”. Multi-depth annual average climatologies retrieved from the World Ocean Atlas were used to contrast environmental conditions at presence vs absence points. For each pOTU, epipelagic (0-200 m) habitat suitability estimates were extrapolated at the level of the global ocean using an ensemble forecasting approach based on four different algorithms (Generalized Linear Models, Random Forest, Artificial Neural Network, and Boosted Regression Trees). For further details about the methodology, please refer to Huber et al. 2024 (submitted in April 2024).

**3.- DATASET DESCRIPTION**

Data type: pOTU habitat suitability maps were calibrated using 2,394 eDNA samples of the pico-size fraction (i.e. 0.2 to 5 μm), retrieved from the EukBank database. Picozoa occurrences were available for 3 different depth levels: epipelagic (0-200 m), mesopelagic (200-1000m), bathypelagic (>1000m). For further details about the dataset, please refer to Huber et al. 2024 (submitted in April 2024).

**Latitude/Longitude format:** EPSG:4326

**Geographic area covered by the dataset:** Global Ocean

**Depth range covered by the dataset:** raw data cover a depth range of 0 – 6015, while maps were extrapolated only for the epipelagic zone (0-200 m) due to lack of sufficient data coverage for deeper layers.

**Time period covered by the dataset:** 2008-2017

**Dataset format:** NetCDF

**Date of dataset creation:**

**Raw dataset repository:**

**4.- MAIN VARIABLE DESCRIPTION**

1 : Mean HSI: average habitat suitability value across all successful models (AUC>0.7)

2 : Min HSI: minimum habitat suitability value across all successful models (AUC>0.7)

3 : Max HSI: maximum habitat suitability value across all successful models (AUC>0.7)

4 : SD: standard deviation of habitat suitability values across all successful models (AUC>0.7)

5 : CV coefficient of variation of habitat suitability values across all successful models (AUC>0.7)

**5.- LINKS**

Article DOI to be added when available

**6.- CONTRIBUTORS**

**Name**

Daniele De Angelis

**email**

deangelis.daniele @yahoo.it

**Position**

Postdoctoral researcher

**Organization**

Sapienza University of Rome

**Location**

Rome, Italy

**Voice**

None

**Fax**

None

**Name**

Paula Huber

**email**

None

**Position**

None

**Organization**

None

**Location**

None

**Voice**

None

**Fax**

None

**Name**

Hugo Sarmento

**email**

None

**Position**

None

**Organization**

None

**Location**

None

**Voice**

None

**Fax**

None

**Name**

Luigi Maiorano

**email**

None

**Position**

None

**Organization**

None

**Location**

None

**Voice**

None

**Fax**

None