**TITLE: AtlantECO [WP2] – Traditional microscopy dataset –** **Pteropoda abundance and biomass concentration data**

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

**1.- INTRODUCTION**

This dataset contains **841 239** georeferenced abundance and biomass concentration records of **56** accepted scientific names of Pteropoda of various taxonomic levels. This dataset is a compilation of the following six main global and regional datasets that reported abundances of marine planktonic pteropods:

* The Coastal & Oceanic Plankton Ecology, Production & Observation Database (NMFS-COPEPOD, O’Brien, 2014) from the National Oceanic and Atmospheric Administration - <https://www.st.nmfs.noaa.gov/copepod/atlas/html/taxatlas_4262500.html>
* The North Atlantic and North Pacific Continuous Plankton recorder (NANP-CPR) survey (Johns D & Broughton D, 2019) - <https://doi.org/10.17031/1629>
* The Southern Ocean CPR (SO-CPR) survey (Hosie, 2021) - doi:10.26179/ksds-s610
* The Australian CPR (AusCPR) survey (AusCPR) - <https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/c1344e70-480e-0993-e044-00144f7bc0f4>
* The Atlantic Meridional Transect (AMT) #24 (Burridge et al., 2016) and #27 (Peijnenburg, pers. Comm.)
* Unpublished calcifying pteropod counts from Ralf Scheibel (Schiebel, pers. comm.).

**2.- METHODOLOGY USED**

We compiled in situ pteropod abundance observations from large scale monitoring programs and existing data compilation programs to calculate total global pteropod biomass and contribution to carbonate fluxes. The main data sources included were the Southern Ocean Continuous Plankton Recorder (SO-CPR) (Hosie, 2021), the Australian CPR (Aus-CPR) (IMOS, 2022), the North Atlantic and North Pacific CPR (NA-NP CPR) (Johns, 2021) and the Coastal and Oceanic Plankton Ecology, Production and Observation Database (COPEPOD) (O’Brien, 2014). Additionally, we included data from the Tara-Oceans program (Brandao et al., 2021), the Atlantic Meridional Transect (AMT24) (Burridge et al., 2016) and AMT27 (Peijnenburg, pers. comm.) and unpublished sampling data (Schiebel, pers. comm.). We matched all taxonomic information against the list of accepted taxon names of the World Register of Marine Species (WoRMS) to harmonize all classifications across datasets and correct for potential deprecated scientific species names (Horton et al., 2017). Observations lacking complete sampling metadata (date, depth, location, abundance value) were removed, together with observations of body parts and plankton observations that did not correspond to the taxa of interest (removal of 580592 points). Additionally, pteropod abundance values from the Ecosystem Monitoring - Ships Of Opportunity surveys (EcoMon-SOOP) in the Gulf of Maine from the COPEPOD dataset were corrected by dividing them by a factor of 100 as the units in the original dataset had been reported erroneously. Biomass conversions were conducted using morphology-based conversion factors from Bednarsek et al., (2012) as detailed in Knecht et al., (in prep.; manuscript available upon demand).

The total pteropod dataset contains 841 239 data points at 309 921 individual locations, collected at a mean sampling depth (± sd) of 38.15 ± 190.89m over the period of 1938 - 2021 (mean ± sd = 2001.25 ± 15.23). Abundances range from 0 to 1066.67 ind.m-3, with a mean (±sd) of 4.38 ± 79.86 ind.m-3. The median abundance (0.00 ind /m3) is low due to the CPR data sets which make up 91.15% of the data, and contain 92.06% absence observations. 50.19% of the data is resolved only on an order level, whereas 24.03% of the observations are species-resolved and 22.41% resolved on a genus level. The dataset contains observations on 33 species out of 165 currently recognized pteropod species (Peijnenburg et al., 2020). The largest contributions to total abundance stem from Limacina helicina (47.7% of the total species-resolved abundance), Heliconoides inflatus (26.7%), and Limacina helicina (10.0%).

The main R packages used to implement this dataset were: ‘tidyverse’ (Wickham et al., 2019), ‘reshape2’ version 1.4.4 (Wickham, 2007), ‘marmap’ version 1.0.6 (Pante & Simon-Bouhet, 2013), ‘lubridate’ version 1.8.0 (Grolemund, 2011), ‘raster’ version 3.5-15 (Hijmans, 2022), and ‘worms’ version 0.2.2 (Holstein, 2018).

**3.- DATASET DESCRIPTION**

**Data type:** Abundances converted to biomass concentrations.

**Latitude/Longitude format:** WGS 84 (-180°E/+180°E).

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

**Depth range covered by the dataset:** From 0m to 4911m.

**Time period covered by the dataset:** From 09-07-1938 to 31-01-2021.

**Dataset format:** .csv file withsemicolon-delimited columns.

**Date of dataset creation:** 14/07/2022.

**Raw dataset repository:** AtlantECO’s GeoNode (<https://atlanteco-geonode.eu/>).

**4.- MAIN VARIABLE DESCRIPTION**

MeasurementTypeID: Has not been defined within AtlantECO

MeasurementValue: Organisms concentration (i.e., abundance) in ind.m-3

MeasurementID: Has not been defined within AtlantECO

occurrenceID: Combination of decimalLatitude, decimalLongitude, Day, Month, Year, Depth, MinDepth, MaxDepth, ScientificName, MeasurementValue.

**5.- DATA OVERVIEW**

**

**6.- CONTRIBUTORS**

- Nielja Knecht (nielja.knecht@gess.ethz.ch), ETH Zürich, Switzerland.

- Fabio Benedetti (fabio.benedetti@usys.ethz.ch), ETH Zürich, Switzerland.

- Urs Hofmann-Elizondo (urs.hofmann@usys.ethz.ch), ETH Zürich, Switzerland..

- Nina Bednarsek (nina.bednarsek@gmail.com), MBS-NIB, Slovenia.

- Katja Peijnenburg (katja.peijnenburg@naturalis.nl), NATURALIS, The Netherlands.

- Catharina de Weerd (cathdeweerd@gmail.com), NATURALIS, The Netherlands.

- Ralf Schiebel (ralf.schiebel@mpic.de), MPI, Germany.

- Meike Vogt (meike.vogt@usys.ethz.ch), ETH Zürich, Switzerland.