**Data extraction DREAM project**

**Extraction performed by Pierre Hélaouët**

**03/02/2025.**

**DOI: 10.17031/67a0958b49501**

**Web page:** **https://doi.mba.ac.uk/data/3428/**

**Note:** As always, this dataset has been carefully built and checked accordingly. However, it is the user’s responsibility to perform his own verifications.

**Quick description of the dataset**

**1 – The dataset contains 14 files:**

1. “CPR\_Data\_DREAMproject\_03022025.docx”: This document
2. “CPR\_ ControlMap\_DREAMproject\_03022025.png”:

Map representing the selected samples from January 1958 to December 2021 (252890 samples)

1. “CPR\_DREAMproject\_Data\_Dinophysis\_03022025.csv”: Abundance data for taxa “Dinophysis spp. Total” (CPR ID = 251) and all selected samples in the selected area (30°N to 80°N, -80°E to 20°E).

Rows: All samples for the selected area (252890 samples).

Column 1: Unique sample id. For instance: “240B--27” corresponds to the 27th sample for the 240th transect on the B route.

Columns from 2 to 8: Spatio-temporal coordinates for each sample.

Columns from 9: Abundance data.

Note 1: We may notice very small values (10^-10) or any other number with a very small fraction. Sometimes, our analysts can identify the presence of a specific taxa but are unable to quantify it. In that case, they report the taxa as “present”. This is hard-coded in our database as a very small value (10^-10) for statistical reasons.

1. “CPR\_DREAMproject\_Data\_Prorocentrum\_03022025.csv”: Abundance data for taxa “Prorocentrum spp. Total” (CPR ID = 259) and all selected samples in the selected area (30°N to 80°N, -80°E to 20°E).

Rows: All samples for the selected area (252890 samples).

Column 1: Unique sample id. For instance: “240B--27” corresponds to the 27th sample for the 240th transect on the B route.

Columns from 2 to 8: Spatio-temporal coordinates for each sample.

Columns from 9: Abundance data.

Note 1: We may notice very small values (10^-10) or any other number with a very small fraction. Sometimes, our analysts can identify the presence of a specific taxa but are unable to quantify it. In that case, they report the taxa as “present”. This is hard-coded in our database as a very small value (10^-10) for statistical reasons.