**Data extraction ICES WGPME WGZE project**

**Extraction performed by Pierre Hélaouët**

**30/10/2024.**

**DOI: 6720c8cd2d827**

Note 1: For each variable, results were checked against data from the previous data extract called “Pierre\_ICES\_Update\_Feb2019.xlsx”, send by Todd OBRIEN on 21/10/2024. Results have minor differences which can be explained:

1. Since the previous extract, the CPR database evolved and now contains “presence data” hardcoded as the value 10-10 (see Note 3). Previously, those data were ignored and coded as 0. It is the responsibility of the user to keep the presence data or set them to 0.
2. Since the previous extract, the Date of Routine Identification (DRI) for “Tintinnida Total“ has been revaluate from 1946 to 1993. Which means that any value for that taxa found before January 1993 is now set to a missing value (i.e., we were not looking for it).

**The dataset contains 2 files:**

1. “CPR\_Data\_ICES\_WGPME\_WGZE\_30102024.docx”: This document
2. “CPR\_Data\_ICES\_WGPME\_WGZE\_30102024.csv”:

For each standard areas (41), the file contains the averaged values of a selected number of planktonic groups (6) for each month (12) of each year from January 1958 to December 2021 (64).

For each row (31488):

1. Column 1: “StdArea\_Nber”: Standard areas number (from 1 to 41).
2. Column 2: “StdArea\_Name”: Standard areas name (from A1 to F10).
3. Column 3: “Year”: Corresponding year (from 1958 tom 2021).
4. Column 4: “Month”: Corresponding month (from 1 tom 12).
5. Column 5: “GroupCount”: Number of samples used to calculate the average.
6. Column 6: “mean\_Diat”: Average values of the sum of all diatoms taxa found in each selected sample (114 taxa).
7. Column 7: “mean\_Dino”: Average values of the sum of all dinoflagellates taxa found in each selected sample (99 taxa).
8. Column 8: “mean\_LCop”: Average values of the sum of all large Copepods (i.e., >2mm, aka “Eyecount”) taxa found in each selected sample (180 taxa).
9. Column 9: “mean\_SCop”: Average values for taxa “total copepods”, CPR ID = 13 (i.e., <=2mm, aka “Traverse”), Aphia ID = 1080.
10. Column 10: “mean\_Tintin”: Average values for taxa “Tintinnida Total”, CPR ID = 10671 (i.e., <=2mm, aka “Traverse”), Aphia ID = 732976.
11. Column 11: “mean\_PCI”: Average values for the “Phytoplankton Colour Index” or PCI. Index based on a semi-logarithmic scale corresponding to the colour of the CPR silk: 0 (no colour), 1 (very pale green), 2 (pale green) or 6.5 (green). For more details, see “Batten et al. 2003. Progress in Oceanography, 58, 193-215.”

Note 2: For groups constituted by several taxa (e.g., Diatoms). First, for each standard area, each year, and each month, all samples corresponding to the targeted taxa (e.g., Diatoms = 114 taxa) are identified. Then, for each sample, a sum of the abundances of all selected taxa is calculated. Finally, an average of all values (i.e., one per sample) is calculated.

Note 3: 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.