

## Issues identified with ICES data on phytoplankton and zooplankton by Germany

### Zooplankton:

At present, no relevant data can be downloaded from the ICES web data portal. This data lacks important information necessary to process and analyze the zooplankton stock composition (e.g., stage-specific, sex-specific data). Metadata associated with this data contains a number of ICES codes (e.g. the cruises from which the data originates, stations etc.) that cannot be assessed or searched on the ICES pages.

Data is, however, available on request. This data does contain stage-specific, sex-specific data. The problems regarding metadata prevail. A short quality check revealed that important information is missing in some data, e.g., the depth strata for the zooplankton catches; without this information the data is useless. Apparently there is no data quality check (either at ICES or at the supplying data center).

Concerning more recent data Germany has corrected biomass calculations for 2011 to 2018 since they were previously not based on Hernroth-biomasses. All German zooplankton data before 2011 that are at ICES cannot be used for the assessment of the HELCOM indicator „Mean Size and Total Stock“.

### Phytoplankton:

#### A) Issues with AphiaIDs

What is ICES doing with species that do not have an AphiaID and/or that are not in the PEG lists? Are they deleted from the database? This should not happen because it destroys the data. If it is necessary to delete selected taxa the whole sample should be deleted. If older data on taxa have to be synchronized to PEG files, the PEG file of the corresponding year and not the latest PEG file has to be used (for details see specific paragraph below).

If taxa without AphiaID or without reference in a PEG list will be neglected this may concern the following taxa:

- 1.) taxonomic groups like the *Diplopsalis* COMPLEX, *Oblea rotunda* COMPLEX, *Scrippsiella* COMPLEX
- 2.) “Unidentified” groups like Pennales, Unicell etc.

On the other hand (probably to avoid missing AphiaIDs), PEG applied the same AphiaID for species and groups, which is, however, not the same, e.g. *Pseudo-nitzschia delicatissima* and *Pseudo-nitzschia delicatissima* GROUP.

The same problem is with "cf.", which indicates a taxon that is not the same as the real species. Is the same AphiaID used for both in order to assure input into the ICES data bank? If not: Has the "cf."-taxon no Aphia-ID and will therefore be completely excluded from the ICES data bank? Or is a validity flag applied for the "cf."-Taxa?

Concerning alignment of taxa names: If older data from previous years are delivered to ICES, containing species names which are not contained in the recent PEG\_BVOL-file: Are these species rejected? If older data are offered which are related to an older PEG\_BVOL file, ICES should be able to load these older data when relating the data to the corresponding PEG-list of the related year. For example: If data from 2013, containing *Cylindrotheca closterium* (Ehrenberg) Reimann & J. Lewin 1958 (AphiaID 149004) are delivered, they should be related to list PEG\_BVOL2014. However if ICES checks species names only with the list PEG\_BVOL2016, this species is not found and rejected because the new name is *Ceratoneis closterium* Ehrenberg, 1839 (AphiaID 163932). Is this the procedure applied? If "yes", this strategy is inappropriate and spoils the data bank because the data are totally incomplete in this case. All data should be loaded with the corresponding PEG\_BVOL-file.

If in older lists AphiaIDs are lacking, this is no reason to reject species data. The key term is the taxon name.

### **B) Issues concerning the data format**

If ICES is contacted, it is possible to obtain data outputs from the ICES database in an excel-readable CSV file. Nevertheless, there is still a lot of work to do before the data can be used in excel. It would be preferable to be able to download data in a more user-friendly format.

### **C) Mistakes and gaps in the data**

When working with the data from the ICES database and comparing them with our data we found numerous mistakes as follows:

- there are spelling mistakes in the older data, including missing or too many blank spaces and classes have been assigned wrongly
- for most sample not all parameters are complete (abundance, biomass, carbon)
- some sample occur twice in the database but with different biomass data (one of these is probably wrong)
- some samples contain very few species (<5) which indicates that the sample has not completely be counted, such samples should be deleted or at least flagged as non-quantitative
- some institutions report extremely low values (calculation mistake or maybe carbon values were reported as biomasses?, deviations are sometimes larger than a factor 100)
- sometimes extremely high values occur beyond any possible phytoplankton biomass

Data search from July 2017 revealed huge data gaps (but it has to be checked whether some of the problems have been solved meanwhile), e.g.

DK: Danish abundance data have problems with the unit: The same number appears with unit "nr/dm<sup>3</sup>" (column "value") and with unit "nr/l" (column "final\_value"); related to the

unit the abundance numbers should differ by a factor of 1000. Biomass data frequently lacking. ES: 1993-2000: data from stations H1, H2 and J1. No data between 2000 and 2010.

FI: 1979-1994. Biomass data from IMRF, but are all too low (could be carbon data?). Data from 1998-2014 ok.

LV: Older data only from 1993 and 1995. Good data series for Gulf of Finland only from 2002. From 2011 to 2015 only single data.

LT: Data from 1991 are much too low. For 1992 data look better but for May 1994 to October 1995 data are again too low. No data from 1996 to 2006.

PL: data only from 1979-1995, 2006-2015.

RU: data only from 1985 to 1986.

SE: most biomass data 1979 to 2004 are much too low. Data from 2008 onwards look ok.

DE: Almost every dataset from 1980, 1981 and 1983 has two identical values for abiotic data and extreme differences in biomass. Very low data from IFOG (up to 100 times too low). German data for the years 1979, 1984, 1985, 1990-2009 are missing.

More specific examples comparing original German data and ICES data:

Problem with *Dictyocha speculum* in sample from 27.10.1998, Stat. TF0360: *Dictyocha speculum* has a biomass of 32.09 µg/l but should have a biomass of 90 µg/l. The same problem occurs in the samples of October 1997, October 2000 etc. Maybe it is caused by problems of converting *Distephanus speculum* to *Dictyocha speculum*?

„*Protoperidinium others*“ on 20.3.1999, Stat. TF0113 should be 384.4 µg/L according to protocol, but in ICES data it is 714.63.

*Thalassiosira baltica* on 22.3.2000, Stat. TF0012 should be 268 µg/L (size class 1+2) according to protocol, but in ICES data it is 393,83.

*Rhizosolenia setigera* on 12.2.2011, Station TF0046 („no named“) biomass of 3.14, but should be 45.14.

LUNG-Data for 2012 are much too low.

More problems might be identified with some additional effort, but perhaps some of the problems are solved meanwhile.