ADDENDUM 2

SUBSETS IN THE UGENT DATABASE

&

SAMPLING TECHNIQUES OF THE MANUELA AND UGENT DATA

ADDENDUM 2

SUBSETS IN THE UGENT DATABASE & SAMPLING TECHNIQUES OF THE MANUELA AND UGENT DATA

The UGent database consists of historical data collected at the Marine Biology Section at Ghent University. It consists of 22 subsets with data collected in the framework of PhD research, BSc dissertations and funded research projects. In the framework of this research hardcopy data were scanned and compiled to a database together with the digital data. An overview of these 22 subsets is given in table A2.1.

Building a consistent database from hard copy data and digital sources is often labour intensive. Firstly, the hardcopy data need to be scanned and after applying optical character recognition the data needs to be corrected and double checked for potential errors. This data is then transformed in a consistent database. This involves different quality checks. The list below shows a brief overview concerning the UGent database:

- Start with designing a hierarchical database which can cope with the different types of data collected for different research topics:
    - Tables: metadata, stations, samples, slices, species data, biometrics, abiotic data, species list;
    - Add the data in a consistent way: Check referential integrity in the database: each species observation should be linked to one slice, each slice should be linked to just one sample, each sample belongs to only one station, each station belongs to one researcher in the database;
    - Check for duplicate entries: identify overlap between data from PhD, MSc thesis and projects;
    - Add meta data about the data supplier and the topic of the research;
    - Add information about each field of the table in the table description;
- Check station data
    - Identify the coordinate system of the supplied data (e.g. ED50, WGS85);
    - Identify faulty coordinates e.g. N>90°, E>180°, swapping of E and N;
    - Identify missing data e.g. 0°N, 0°E is most probably incorrect;
    - In case of marine data: coordinates should be situated at sea;
    - Spatial references which are plausible, but incorrectly described in the original source, are often hard to trace (e.g. in the UGent database there was one station which always represented an outlier in the analyses. Eventually, while double checking the original literature source, it was clear that the coordinates in the table and the coordinates on the map represented different coordinates. The map showed the correct position of the station. By replacing the coordinates by those represented in the map, the data were no longer outliers.)

- Add sample data:
    - Date of sampling (if known: hour of sampling);
    - Sample gear, sample size, subsample size;
    - Identify replicate samples;
    - Add total density of the core sample;
- Add slice data:
    - In order to construct a consistent database identify the whole core as a slice if no separate slices where taken;
    - Add total density of the slices;
- Add species data
    - Check validity of nomenclatural and taxonomic classification
        - check validity of names
        - identify synonyms (check literature sources and taxonomical websites such as NeMys, ERMS, WoRMS, …)
        - identify misspellings (find similar names and identify the correct spelling)
        - standardise records with an uncertainty level (e.g. aff., n.sp.1, …) and identify the closest most specific taxonomical level. However, keep the original data. It often holds additional information.
        - identify misidentifications (hard, if not impossible, to trace)
    - Check species numbers
        - Include as much as possible the original data:
            - Species counts are preferred to abundances if the total abundance in the sample is known;
            - Abundances are preferred to relative abundances;
            - Relative abundances are preferred to presence/absence.
        - Number of counts should never be larger than the total species density in the sample;
        - The sum of the abundances per species should never be larger than the total density of the sample;
        - The relative abundances should never add up to more than 100%;
- Add environmental data
    - Add information on how the environmental data was measured. Check if the different methods result in different accuracies (often hard to trace);
    - Specify matrix where the environmental data is measured: water, sediment (e.g. chl *a*);
    - Add a unit to each variable;
    - Standardize units as much as possible (e.g. µg/l ~ g/m³);
    - Describe units as specifically as possible (e.g. weight% or volume% for sediment fractions);

    - Identify improbable or impossible values of the environmental variables (e.g. sediment fractions add up to more than 100%; sort from largest to smallest values and identify extreme outliers);
    - Write percentages in the same way e.g. 1%=0.01, the value may be entered as 1 or as 0.01 in the database. Specify in the description of the table;
    - Add environmental data extracted from maps, but give them a different label;
- Add additional data to the database:
    - Feeding type for each species;
    - Taxonomical tree: family, order, class;
    - Information known from some data sources but not specified in others (such as sampling design, species counts, total densities, …) was traced back in original descriptions as much as possible.
- When in doubt always contact the data supplier and ask feedback!

Data provider	Data source	Description	Time frame
Ann Vanreusel	Hardcopy	PhD: Ecology of the free-living marine nematodes from the Voordelta (Southern Bight of the North Sea).	1988-1989
Carlo Heip	Digital	Nematodes from the NSBS. Data collected in the framework of the 1986 North Sea Benthos Survey, an activity of the Benthos Ecology Working Group of ICES.	1986
Chen Guotong	Hardcopy	BSc Thesis: Study of the meiobenthos in the Southern Bight of the North Sea and its use in ecological monitoring.	1987
Gonda Bisschop	Hardcopy	BSc Thesis: Study of the nematode fauna of the North Sea and the mouth of the Western Scheldt estuary.	1976-1977
Jan Vanaverbeke	Digital	Nematodes from station 330: structural and functional biodiversity on the Belgian Continental Shelf (TROPHOS project).	1999
Jan Vanaverbeke	Digital	Study of the ecological effects of sand extraction on the Kwintebank: evaluation of past extraction effects. (SPEEK project).	2003-2006
Jian Li	Hardcopy	The temporal variability of free-living nematodes in a brackish tidal flat of the Western Scheldt with emphasis on the use of an ecological model.	1992-1993
Jyotsna Sharma	Hardcopy	PhD: A study of the nematode fauna of three estuaries in the Netherlands.	1985
Maaike Steyaert	Digital	Meiobenthos at the stations 115, 702, 790 on the Belgian Continental Shelf (IMPULS-project).	1994
Maaike Steyaert	Digital	Spatial heterogeneity of nematodes on an intertidal flat in the Western Scheldt Estuary. (Ecoflat project).	1997
Maaike Steyaert	Digital	Nematode data of a station at the German Bight.	2002
Maaike Steyaert	Digital	Meiobenthos station 115bis - benthic-pelagic coupling (TROPHOS/PODO-I work-database I 23/01/2004):	2004
Maarten Raes	Digital	PhD: An ecological and taxonomical study of the free-living marine nematodes associated with cold-water and tropical coral structures.	2006

Data provider	Data source	Description	Time frame
Magda Vincx	Hardcopy	PhD: Free-living marine nematodes from the Southern Bight of the North Sea.	1987
Matthew Lammertyn	Digital	BSc Thesis: Live on and around ship wrecks on the Belgian Continental Shelf (BEWREMABI project).	2005
Preben Jensen	Hardcopy	BSc Thesis: Nematode fauna on the silty and sandy sea floor in the southern North Sea (I.C.W.B.).	1974
Regine Vandenberghe	Hardcopy	BSc Thesis: The meiobenthos in a dumping site at the Southern Bight of the North Sea, with emphasis on the free-living marine nematodes.	1987
Sandra Vanhove	Digital	Data collected in the framework of the ANDEEP 2 project.	2002
Sandra Vanhove	Digital	Data collected in the framework of the LAMPOS-ANDEEP-project.	2002
Sandra Vanhove	Digital	Data collected in the framework of the EPOS-leg3-project.	1989
Saskia Van Gaever	Digital	PhD: Biodiversity, distribution patterns and trophic position of meiobenthos associated with reduced environments at continental margins.	2008
Tom Gheskiere	Digital	PhD: Nematode assemblages from European sandy beaches diversity, zonation patterns and tourist impacts.	2005
Zhang Derong	Hardcopy	BSc Thesis: Evaluation of the meiofauna and nematode community at a TiO_2-dumping site after recovery.	1995

Table A2.1 Data providers, source, description and time frame of the 22 subsets in the UGent database.

Data provider	Sample device	Sample area (cm²)
Magda Vincx (Chen Guotong)	Box corer	10
Maaike Steyaert & Li Jian	Cores	10
Magda Vincx (Jyotsna Sharma)	Divers and Reineck box corer	10
Jan Vanaverbeke & Magda Vincx	Reineck box corer	10
Ann Vanreusel	Van Veen or box corer	10
Michaela Schratzberger	Bowers & Connelly multi corer	23.76
John Lambshead	Core	3.8
Tim Ferrero	50 ml syringe corer	5.31
Michaela Schratzberger	Bulk	unknown
Magda Vincx	Van Veen, box corer or diver	7 to 18.4
Andrea McEvoy	Craib cores & Day grab	unknown

Table A2.2. Sampling techniques of the MANUELA data used in Chapter 2.

Data provider	Sample device	Sample area (cm²)
Preben Jensen - BSc Thesis	Van Veen	11
Jan Vanaverbeke - Trophos	Reineck	10
Maaike Steyaert - Trophos	Reineck	10
Magda Vincx – PhD	Diver, Reineck or Van Veen	7 to 18.4 cm²
Jan Vanaverbeke - Speek	Reineck	10
Chen Guotong - BSc Thesis	Box corer	10

Table A2.3. Sampling techniques of the UGent data used in Chapter 3.

Data provider	Sample device	Sample area (cm²)
Ann Vanreusel (PhD)	Box corer	10
Ann Vanreusel (PhD)	Van Veen	10
Preben Jensen (BSc Thesis)	Van Veen	11
Jan Vanaverbeke (Trophos)	Reineck	10
Maaike Steyaert (Impuls)	Box corer	10
Maaike Steyaert (Trophos)	Reineck	10
Magda Vincx (PhD)	Box corer	10
Magda Vincx (PhD)	Diver	10
Magda Vincx (PhD)	Reineck	7 to 18.4 cm²
Magda Vincx (PhD)	Van Veen	7 to 18.4 cm²
Jan Vanaverbeke (Speek)	Reineck	10
Chen Guotong (BSc Thesis)	Box corer	10

Table A2.4. Sampling techniques of the UGent data used in Chapter 4.

Data provider	Sample device	Sample area (cm²)
Ann Vanreusel (PhD)	Box corer	10
Ann Vanreusel (PhD)	Van Veen	10
Regine Vandenberghe	Box corer	10
Jan Vanaverbeke (Trophos)	Reineck	10
Maaike Steyaert (Impuls)	Box corer	10
Maaike Steyaert (Trophos)	Reineck	10
Magda Vincx (PhD)	Diver	10
Magda Vincx (PhD)	Reineck	7 to 18.4 cm²
Magda Vincx (PhD)	Van Veen	7 to 18.4 cm²
Magda Vincx (PhD)	Box corer	10
Sharma Jyotsna (PhD)	Reineck	10
Jan Vanaverbeke (Speek)	Reineck	10
Zhang Derong (BSc Thesis)	Box corer	10
Chen Guotong (BSc Thesis)	Box corer	10

Table A2.4. Sampling techniques of the UGent data used in Chapter 5 and 6.