

Interactive comment on “CASCADE – The Circum-Arctic Sediment CARbon DatabasE” by Jannik Martens et al.

Gerrit Müller (Referee)

g.muller@uu.nl

Received and published: 27 January 2021

General Comment

The authors generated a comprehensible, freely accessible dataset on organic carbon concentration, its isotopic composition, nitrogen concentration and terrigenous biomarkers in circum-Arctic marine sediments. This is an original and useful compilation, as it makes data published in different sources easily accessible, previously unpublished data available and even complements existing data with gap-filling measurements and spatial interpolation using GIS. Such data is needed for integrative and large-scale assessment of biogeochemical cycles, especially in regions as sensitive to changing climate as the Arctic. First insights are deduced and visualized, identifying

C1

spatial variation in organic carbon concentration and sources, which gives the reader an impression of how to use and interpret this data. The provided dataset is well organized, curated, (re)traceable and citable via DOI, despite a few apparent transcription errors (which are common in these types of manual compilations) and minor comments on comprehensibility (see Dataset comments). The paper itself describes data derivation and classification in a reasonable way and with sufficient detail, although (apart from minor comments) the reader-friendliness of the structure, consistency and conciseness could be improved.

Specific Comments

Abstract The Abstract includes all relevant information, but could be more concise. For instance, the beginning (L.30 – 34) can be reformulated more concise and your interpretation presented in L. 46 – 50 may be explained more briefly. Possibly, try to avoid too many adjectives/adverbs and embedded subordinate clauses.

Introduction 1.) In L. 58: '[...] large input of terrestrial organic matter from its large rivers and from coastal erosion, making it both a valuable receptor system for studying large-scale terrestrial carbon remobilization and marine biogeochemistry', consider substitution of one or two 'large'. There is also a semantic error: '[...] making it both a valuable receptor system [...] and marine biogeochemistry.' You may try to relocate 'both'. 2.) References are needed to support your statement, that warming accelerates coastal erosion and river runoff in L. 60 - 61 (although this is senseful of course). Similarly, citations should be added in L. 69-70 (arctic warming as a tipping point in the climate system) and L. 75 – 79 (global and arctic shelf area portions). In L. 89 – 90, where you mention 'Key progress', the (or some) relevant articles should be cited. If that refers to the references in the following explanatory sentences, please make that more clear, e.g. by inserting ':'. 3.) In L. 108 – 109, you state that there is an initial focus on terrigenous organic matter, but not why. Is that because of data availability or because of the database applications you had in mind? Please give the reasoning here.

C2

Data collection and methods 1.) Section 2.2 Georeferencing and sampling is a bit confusing (e.g., what do you mean by 'core part?'), which is clarified later in section 2.4 Database parameters. So maybe consider explaining the parameters and structure before you give details on how these parameters are acquired. 2.) L. 151 states that you used '[. . .] the year of the earliest published record [. . .]' when the sampling date was not available. This needs to be visible in the data tables, for full comprehensibility and because publication and sampling year may differ in fact. Is more detailed sampling time information available? Seasonality may play a role in surface sediments. 3.) In section 2.3 you could possibly provide the number of sediment cores available for each of the scales (Centennial, Millennial and Glacial cycles scale) and provide a reasoning for this separation (i.e., how did you choose length range and corresponding time-scale?) 4.) In L. 206 – 209, it is stated that for some samples, variables are from different references. Is that exactly the same sample (or a split)? How do you assure this? Is it explicitly stated by the references? 5.) The detailed description of the included parameters in section 2.4 Database parameters is (partially) repeated in section 3.1 Data set inventory, where you basically make a similar detailed description, but with numbers of samples. 6.) Section 2.6 Data source and quality. In this section you describe 'The quality criteria for data to be included [. . .]'. These appear well chosen at the first glance, but later in the evaluation of the data quality, it becomes obvious that not all data fulfill these criteria (e.g. 'For $\delta^{13}\text{C}$ -OC, in 66 % of the cases IRMS coupled to EA was given [. . .]' (L. 333 -334) and the rest?). Therefore, you may not call these selection criteria, but quality assurance criteria or similar. At least stay consistent with this. 7.) In section 2.7.1, L. 259 – 260, you give the precision of the gap-filling measurements, but neither mention how this was acquired (multiple measurements?), nor include a statement regarding accuracy. 8.) The constants used for conversion of ^{14}C data should be referenced (L. 282 – 284).

Results and Discussion 1.) Section 3.1: Partial repetition of section 2.4. See comments in previous section. 2.) Section 3.3: Does your remark in L. 332-334, '[. . .] CASCADE provided detailed information [. . .] or cited references or cruises

C3

that contained this information', mean that it may be up to the user to look up methods for some samples in the original reference? If yes, why didn't you include this information for all samples then? If no, please try to reformulate to avoid confusion. 3.) Which external data sources are available that can be interoperated with CASCADE? It may be instructive to compare marine sediment data to river input, water column chemistry, marine production, coastal and soil erosion or the non-organic carbon part of marine sediment. At least for some of these parameters, databases exist (e.g. GLORICH <https://doi.org/10.1594/PANGAEA.902360> or db seadbeds <https://instaar.colorado.edu/~jenkinsc/dbseabed/>). You may stimulate users by providing some advice here. 4.) You may possibly want to insert a disclaimer in the end (<https://www.earth-system-science-data.net/submission.html#manuscriptcomposition>).

Dataset 1.) CASCADE is very well structured, comprehensible and user friendly. However, an indication for what type of time-reference is given should accompany that valuable information. In the current version, a user cannot know if a sampling time refers to the real sampling year or to the earliest publication year from the same sample/core, which may differ by some years. 2.) The dataset has proven robust against detailed inspections and logical tests. In spite, there is a little transcription (?) error, identified by the histogram of '% OC' in 'CASCADEcoresv1', indicating an impossible organic carbon concentration of ~ 120 %, which probably propagated into the following calculations (C/N ratios and normalization to organic carbon concentration). Please check your calculations, possible errors in unit conversion and transcription errors (digit typo?) and correct this. 3.) A table translating abbreviations used in the database (e.g. IRMS) and units should accompany the data for clarity and convenience, despite this information is included in the article. 4.) If possible, please provide the interpolated grid also as datafile to complement the rather impractical tiff-file.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-401>, 2020.

C4