

1 Dear Editor,

2

3 Thank you very much for relaying these reviews to us. We are very encouraged that the
4 reviewers are so positive, and they see the value of this new database. We have incorporated
5 all revisions, and we greatly appreciate the time the reviewers have taken to provide detailed
6 and constructive comments. The major improvement, suggested by both reviewers #1 and #2,
7 was to provide additional details with respect to the data collection process and data
8 conversions (e.g., of the coordinate systems). We agree, and we have now both included a short
9 summary in the main text and have provided more details in a dedicated section in the
10 Supplemental Information. Please see the point-by-point replies below for details.

11

12 Again, we greatly appreciate the detailed feedback, and we hope that the revised manuscript,
13 and the database described therein, will provide a foundation for further research in marine
14 carbon and in the wider earth sciences.

15

16 Many thanks in advance for considering this revised manuscript,

17

18 On behalf of all co-authors,

19

20 Tessa van der Voort

21

22 [Point-by-point replies referee #1](#)

23

24 I enjoyed reading the manuscript by T.S. van der Voort et al. entitled “MOSAIC (Modern
25 Sediment Archive and Inventory of Carbon): A (radio)carbon-centric database for seafloor
26 surficial sediments”. The need for a surficial sediment database for organic carbon and
27 radiocarbon is well justified in the text. I would like to see the MOSAIC database established,
28 and as an indication of my support and approval, I am likely to contribute most of my
29 radiocarbon and organic carbon data to such a program. I like the idea of using open-source
30 software and making this focused database convenient to the scientific community and user
31 friendly.

32

33 Thank you for this! We look forward to incorporating your data. That's how we can help
34 MOSAIC grow!

35

36 The only concern I have about MOSAIC, as it strives to become accepted as a global database,
37 is that I wasn't certain as to which 200 papers were used to establish the initial database. The
38 authors make generalizations regarding C-13 and C-14 data in the discussion section, but the
39 rigor of these generalizations depends on which 200 papers were used to establish the data
40 base. Were these primarily papers written by Tim Eglinton's research group or was a broader
41 approach used in the selection of the organic carbon and radiocarbon data?

42

43 Thank you for this suggestion. We have included a short description of how the papers were
44 selected in the main text, and a detailed description in the SI. A broader approach was used to
45 select the organic carbon and radiocarbon data, building upon an initial synthesis effort
46 (Griffith et al., 2010) where papers (not only those from the Eglinton research group)
47 containing TOC and ^{14}C data from several margins was used as a starting point. This was then
48 augmented by a subset of papers containing similar information (i.e., sediment TOC and ^{14}C
49 data) selected by the senior author (Eglinton), which were used to familiarize the researcher
50 involved (Usman) in extraction and assessment of relevant data. Then, additional papers were
51 sought out using google scholar, using search terms such as "TOC in surficial sediments",
52 "organic carbon in surficial sediments" and " ^{14}C /Radiocarbon in surficial sediments" and
53 relevant data were ingested into MOSAIC. Where possible, references in found papers were
54 followed up on to access more or original datasets. This resulted in over 200 papers.

55

56 There is reasonable global coverage of continental margin sedimentation in the MOSAIC data,
57 but there are some obvious holes in the database, such as the continental margin sediments
58 surrounding the Antarctic Peninsula (where there has been substantial radiocarbon data
59 published in the past several years).

60

61 We are aware that our data search and ingestion process has far has not been exhaustive, due
62 to the limitation of personnel dedicated to this activity. The database is built to be dynamic and
63 can continually absorb more datasets, and it is intended to growth both through further
64 combining of the literature for additional data, through on-going acquisition of new data, and
65 as scientists become aware of MOSAIC and contribute their own data. It has been noted for

66 the future updates. MOSAIC makes preliminary generalizations based on the 200 papers, but
67 this only represents a fraction of the available literature, particularly given that our search terms
68 may have missed numerous contributions. At this point, however, with this data collection and
69 the developed digital infrastructure, we felt it timely to prepare an initial publication. A newly
70 funded post-doc position in the Eglinton Group (starting March 2021) will be dedicated to the
71 MOSAIC database.

72
73

74 The manuscript does describe QA/QC concerns of the radiocarbon and total organic carbon
75 data, but these are primarily from a statistical perspective. Very little is mentioned in the
76 manuscript about analytical concerns, blank issues, and potential contamination during
77 sampling.

78 Thank you for these comments. Indeed, there is a lot of focus on the statistical and automated
79 QA/QC. When the data was collected from the papers, care was taken to take from trusted,
80 peer-reviewed sources. However, we also have to trust the quality of the researchers, labs and
81 peer-review with respect to their data processing and reporting. Whenever it was available, we
82 included reported uncertainties (error values). Due to space limitations these are not included
83 on the website-based version of MOSAIC. However, in the SQL-based database, which is also
84 included, this data can be accessed. We included an introductory guide in the SI on how to use
85 specific database queries. We have added an example that also extracts all error values. For the
86 most common parameters of course, the MOSAIC website is designed to provide a user-
87 friendly, intuitive interface.

88

89 The database currently lists radiocarbon data using a Fraction Modern (Fm) nomenclature, but
90 mentions that Delta14C nomenclature will ultimately be used for the database. If so, I would
91 recommend that the authors include a “Date of Collection” data box in their submission data
92 and website display, so that users can easily go back and forth between Fm nomenclature and
93 Delta14C nomenclature. In fact, I think it would be useful to list the radiocarbon data using
94 both the Fm and Delta 14C formats.

95

96 Thank you for this comment. Regarding, Fm versus Delta¹⁴C, whenever it was possible, the
97 sampling year was known or collected, we converted Fm to Delta¹⁴C and vice versa. For the
98 future data ingestion (also from our fellow researchers), we have a field in the submission

99 excel-sheet that includes the sampling year, so this conversion can be done. On the MOSAIC
100 website (mosaic.ethz.ch), both Fm to Delta¹⁴C data are directly available.

101

102 In addition, I also would recommend that the authors consider adding to the data input table
103 the type of coring device used to collect the marine sediments. There is a big difference between
104 the quality of surficial sediment collected by a multicorer or megacorer as compared to a kasten
105 corer or piston corer. Such information would be useful to a researcher comparing organic
106 carbon or radiocarbon abundances over a basin or region.

107

108 Thank you for this comment. Whenever available, we collected data on the coring device, and
109 this is also included in the database, but for simplicity this information is not included in the
110 web-based interface. We provided a specific example on how to retrieve information on the
111 corer type in the available introductory guide on how to use specific SQL queries.

112

113 I think that the authors make the case that radiocarbon data are the most needed information
114 for continental margin databases. That being said, of the total 8706 data entered into MOSAIC,
115 there are only 709 radiocarbon measurements (as compared to 8688 analyses of Total Organic
116 Carbon). Thus, although radiocarbon may be the primary emphasis of the MOSAIC database,
117 it represents less than 10% of the data entered into the system.

118

119 Thank you for this comment. Due to the high monetary and labor costs, ¹⁴C measurements
120 remain much rarer than TOC measurements, and the proportion of ¹⁴C to TOC data in MOSAIC
121 reflects that found in most papers. Most papers contain around a handful or a dozen ¹⁴C
122 datapoints, so even to attain this value a considerable effort was needed, and we intend to keep
123 adding to it. However, this situation is changing rapidly as the number of accelerator mass
124 spectrometry systems installed around the world has dramatically increased within the past
125 decade. A key priority in the future development of MOSAIC is to incorporate newly reported
126 and recently acquired ¹⁴C data, and this will be a primary focus of a newly funded post-doc
127 position (starting March 2021).

128

129 The MOSAIC database also lists the Calcium Carbonate content and the Silicate (SiO₂) content
130 of the sediments. The text does not reference how these measurements were made or even if
131 the silicate abundances includes biogenic silica with the lithogenic silica content.

132 The focus of data collection activities thus far have been on the abundance and characteristics
133 of organic matter, – however, the relevance of inorganic components with respect to, for
134 example, biogeochemical fluxes and organo-mineral associations is also recognized. Thus far,
135 no distinction has been made between biogenic (opal) or lithogenic silicate sources, but this
136 may be further defined in subsequent iterations of MOSAIC. In the meantime, if researchers
137 want to further explore such parameters, the DOI is provided for each datapoint, and they can
138 easily attain more details about the data with one click.

139

140 Minor Suggestions Reviewer #1

141 The manuscript could have been proofread more thoroughly prior to submission. For example:

142 -On lines 159-161 the words don't comprise a complete sentence.

143 Corrected, thank you

144 -On line 193 add commas on either side of “for example”.

145 Corrected, thank you

146 -On line 276 the text reads “rather that” and it should be “rather than”.

147 Corrected, thank you

148

149

150 -On line 289 “exhibits” should be “exhibit”. Corrected, thank you

151 -On line 327 change “couple” to “couple with”. Corrected, thank you -On line 336 add “of”

152 before “geochemical”. Corrected, thank you -On line 345 change “14C” to “14°C”. Corrected,

153 thank you -On line 363 change “derives” to “was derived”. Corrected, thank you -On line 370

154 change “explain users” to “explain to users”. Corrected, thank you -In Fig. 5 the partial

155 derivative sign is used instead of the small Greek symbol delta. The Greek symbol is used

156 correctly in Fig. 4, but the partial derivative symbol needs to be changed to a lower case delta

157 symbol in Fig. 5. Corrected, thank you

158

159 On lines 128 and 179, the authors should consider not only listing the “mixed-layer depth”, but

160 also include “bioturbation intensity” as a parameter for characterizing the nature of surficial

161 sediments

162

163 We have included this, thank you.

164

165 . On line 293 the text states: “ageing associated with sediment reworking by bottom currents”.
166 The authors should mention bioturbation as well as physical sediment reworking. It is much
167 more likely that continental shelf and continental margin sediments are mixed by bioturbation
168 than by physical reworking.

169

170 Thank you for this comment. In the paper cited here, the focus was on physical reworking of
171 sediments via lateral redistribution, however the reviewer is absolutely right that bioturbation
172 is an extremely important consideration. We have highlighted the importance of bioturbation
173 in line 128. Incorporation and parameters and data related to bioturbation (e.g., sediment mixed
174 layer depth, oxygen penetration depth) will be a focus of the next iteration of MOSAIC.

175

176 On lines 318-322 the text reads: “The latter is particularly pertinent for 14C data and ancillary
177 measurements necessary to broadly apply isotopically-enabled models of organic turnover and
178 burial in sediments (e.g., Griffith et al., 2010) and constrain geographic variability in the age
179 distribution of sedimentary OC . . .”. I suggest that the authors consider adding the following
180 reference after the Griffith et al., 2010 citation:

181

182 Isla and DeMaster, 2018 (GCA, v. 242, 34-50; entitled “Labile organic carbon dynamics in
183 continental shelf sediments after the recent collapse of the Larsen ice shelves off the eastern
184 Antarctic Peninsula: A radiochemical approach”). This paper is a recent example of
185 “isotopically-enabled models of organic turnover”.

186

187 Thank you, we concur, and have included a reference to this informative paper.

188

189 Why do the authors use the word “seafloor” in the title instead “marine”. Using “seafloor” and
190 “sediment” so close to each other seems redundant to me.

191

192 Thank you for the comment. In the title we want to make clear that the focus of this database
193 is on surficial marine sediments (i.e., not the longer cores such as those acquired IOPD cruises
194 for paleoclimatic studies). Thus, we opt to use “seafloor” as we believe it best implies we are
195 discussing surficial ocean bottom sediments.

196

197 In summary, I support publication of the MOSAIC ESSD article after minor concerns,
198 mentioned in the review above, have been addressed by the authors. I encourage the authors to
199 continue their efforts to develop and create these new databases that enable scientists
200 easier/facilitated access to organic carbon and radiocarbon data published in the marine science
201 literature.

202

203 Thank you, we very much appreciate the detailed and constructive comments and the time
204 you've taken to provide them. We also appreciate the sentiment that this contribution will
205 further scientific research on ocean sediments. We have processed the comments and look
206 forward to sharing this work with the scientific community.

207

208 Anonymous Referee #2

209

210 Major comments:

211 In general, I welcome the proposed database and can see its value and utility. However, I do
212 have several points to raise to the authors that should be addressed before publication:

213

214 Thank you for this comment. We've addressed the points that have been raised, and you are
215 very grateful for the time and effort you have put into this review.

216

217 The narrative in the Introduction forms a case for support for the need and uniqueness of the
218 database on the one hand, whilst on the other slips into scientific argument of what
219 could/should be done with the data. Both articulations are reasonable, but confuse the reader
220 somewhat. I suggest toning down the suggestions on what can be done with the data. Overall,
221 the paragraphs starting line 82 and line 120 seem largely redundant. Similarly, i was surprised
222 not to see reference to recent reviews and opinion pieces about sediment carbon (e.g. Snelgrove
223 et al. 2018, TRENDS IN ECOLOGY & EVOLUTION, 33, 96-105; Middelburg, 2018
224 BIOGEOSCIENCES 15, 413-427) to reinforce what we know and what we don't know. these,
225 and other similar summaries should be incorporated into the text.

226

227 Thank you for these comments, we have endeavored to incorporate your suggestions.

228 Indeed, we agree with your suggestion that the focus of this paper lies on presenting the
229 MOSAIC dataset and digital infrastructure. Citations for the mentioned opinion and summary
230 papers (Snelgrove et al., 2018 and Middelburg 2018), which have now been incorporated in
231 order to underline the broader utility of MOSAIC for the scientific community, e.g., by
232 improving the robustness of sedimentary organic carbon turnover estimates and the
233 understanding of organic matter processing in seafloor sediments. We have now also included
234 additional summary papers, such as those by Arndt et al., (2013) and Bianchi (2011) that
235 highlight the need for the type of information residing in MOSAIC.

236

237 Presently the paper is structure that it provides only a cursory glance at the available data by
238 way of illustration of the sorts of information that can be retrieved. We have visualized the
239 data, but it is not intended to be a rigorous assessment or provide in-depth interpretation (e.g.
240 quantifying carbon stocks or using machine learning algorithms to extract spatial patterns as

241 e.g. we have done in respectively Avelar et al., (2017) and van der Voort et al., (2018)). We
242 now emphasize this point, and also point out that we have modelled the structure of this
243 manuscript to follow others in this journal that announce a database and provide examples of
244 data content.

245

246 2. MOSAIC - minor point, but this acronym is a little unfortunate as it matches the MOSAIC
247 expedition in the Arctic (<https://mosaic-expedition.org/>), a significant programme that will
248 have a long legacy in the literature. I suggest altering the acronym to avoid this overlap, and
249 suggest the authors consider using a title rather than an acronym that incorporates the
250 description of the exactly what is in the database.

251

252 Indeed, MOSAIC is an acronym that occurs in other settings, and indeed also for the Arctic
253 expedition. The latter is a field program, while our website (mosaic.ethz.ch), clearly
254 immediately refers to being a database. We believe scientists will be able to readily make this
255 distinction. Furthermore, the capitalization of both abbreviations is different, where the
256 database is all caps (MOSAIC), the expedition has a lower-case “I” (MOSAiC). We believe
257 MOSAIC is an apt name, because we investigate spatial mosaics in geochemical and
258 sedimentological properties on the ocean floor.

259

260 3. Line 146 (and then Line 170)- I see the intention of the database, but how often will it be
261 updated and what data quality controls are in place?

262

263 Thank you for this comment. We aim to announce a quasi-yearly update, with the most up-to-
264 date version mentioned on the website. Starting March 1st, a dedicated post-doc will be fully
265 devoted to this project, with a focus on further ingestion of data (esp. ¹⁴C) and expansion of
266 parameters. Regarding the quality controls, (as mentioned in section 2.3.2), we have an initial
267 auto-check written in Python which checks data and flags unusual data (e.g. TOC values that
268 are <0). We have now added a detailed description of all automated checks in the SI. After the
269 automated check, a member of the ETH MOSAIC Team will manually check the flagged data.
270 This ETH MOSAIC team member will also perform an additional manual check to see if all
271 the data was read in correctly.

272

273 re line 170, how with the new information gel with the older data, and will efforts be made to
274 back fill the missing data?

275

276 MOSAIC has been explicitly designed to be a dynamic database. Data can be added and
277 ingested easily, as described below. As part of these efforts, there will be targeted efforts to
278 “back-fill” missing data, as we continue to uncover previously published work. Indeed, this is
279 one of the defining attributes of MOSAIC – that it has a specific objective to collate and
280 organize data germane to the overall theme of organic matter accumulation on continental
281 margins, instead of serving as a passive repository for data. We are also open to incorporating
282 new variables for future versions (e.g., those relevant to seafloor ecology) if they are brought
283 forward by the community. Thanks to the adaptable SQL framework, this would involve just a
284 few lines of new code.

285

286 New data can be ingested in the provided spreadsheets which have built-in vocabularies (e.g.,
287 for corer types or ocean names). Then, the data in spreadsheets (Microsoft or LibreOffice) will
288 be converted to be SQL-ingestible. This is done by using Python scripts that automatically add
289 unique identifiers to the data and convert the Microsoft Excel or LibreOffice files to csv files
290 which can be ingested in the mySQL environment.

291

292 We aim to continually expand the MOSAIC dataset. For example, a key next goal is to develop
293 carbon inventories of sediments according to the Economic Exclusive Zones (EEZs), and to
294 identify regions where data is particularly sparse. The continuous addition of data (new and
295 old) to MOSAIC will enhance the value for the scientific community.

296

297 More technical details on Quality Control are section 2.3.3.

298

299

300 4. Paragraph starting Line 164 - A very important aspect of any database that has extracted
301 information from the literature is that the search terms and process of selection criteria needs
302 to be repeatable and absolutely clear. This is of fundamental importance and needs to be
303 explicitly stated in the this section with supporting information in the supplementary material.
304 How were the 200 papers found, selected and checked for data? What search engines and
305 search terms (including any refinements) were used, and how were quality controls
306 implemented? How many papers did the initial search yield, and how was the final subset
307 arrived at? When was the database accessed? Does this database contain data from other
308 databases? What downstream processing of the data, or meta-data, was necessary? e.g. were
309 units converted, how was lat and long derived/converted to the same projection, how was a
310 position assigned to biogeographical zones etc? All steps need to be explained. This is an
311 essential area that needs to be articulated in detail to ensure the authority of the data. The
312 authors need to convince the reader that these data are the ones to use. This is probably the
313 most important aspects of my commentary that needs addressing fully. Section 2.1.2 needs
314 significant amendments with a focus on attention to detail.

315

316 Thank you for this comment. We fully concur with this point and have revised and expanded
317 the text accordingly.

318 We have included a brief summary answering the issues you raised in the main text and added
319 a highly detailed section in the supplemental information.

320 To answer your question directly here:

321 **Q1A: How were the 200 papers found**

322

323 The current MOSAIC dataset was initiated by manual mining of an initial subset of
324 peer-reviewed oceanographic papers that contained substantial TO¹⁴C datasets (e.g.,
325 Griffith et al., 2010) from different continental margin systems. This initial dataset was
326 collected by an experienced oceanographer, this papers' senior author (Eglinton) This
327 enabled the collecting researcher (Usman) to be trained in the process of data evaluation
328 and handling.

329

330 MOSAIC was further expanded by extracting data from a broader suite peer-reviewed
331 papers which were found using the search engine Google Scholar, with search terms
332 including “organic carbon in surficial/surface sediments”, “TOC in surficial/surface

333 sediments” and “radiocarbon/¹⁴C in surficial/surface sediments”.”. When appropriate
334 papers were found, references were followed up on to find similar contributions in the
335 region. This yielded several hundred of papers.

336 **Q1B selected?**

337 From the several hundreds of papers, only papers that contained the required parameters
338 for data were retained (i.e., lat. and long. for each TOC or ¹⁴C data point). Furthermore,
339 the papers which focused on sediment dissolved organic carbon or inorganic carbon
340 were excluded given the focus on the solid phase and organic phase, and a priority on
341 surficial sediment data, captured by corers that best preserve the sediment-water
342 interface (i.e., multicorer or box corer.). Other corers are not strongly represented.
343 Furthermore, papers were selected for whom the data was available in tabulated for, i.e.
344 not exclusively in graphical form in order to ensure the quality of extracted data (see
345 next paragraph for more details).

346

347

348 **Q1C and checked for data?**

349 As mentioned, for the older papers, the researcher extracted the data manually, point-
350 by-point from tables or exceptionally from graphs in papers in pdf format. While this
351 process is very laborious, it enables the scientific community to access data which
352 would otherwise potentially be lost in time. On the rare occasions where the sampling
353 locations are presented as dots on a map (without accompanying exact geographical
354 information), the longitudes and latitudes were “hand-traced” and the approximate
355 geographical information were reported. We acknowledge that this process is
356 accompanied with uncertainties, but feel they are acceptable given the value and
357 irreproducibility of the data in older papers.

358 We believe this manual data extraction from older papers has a significant added value,
359 as for normal research projects it would not be feasible to invest this time.

360

361 For many of the more recent papers, the researcher could extract data from csv files or
362 paper SI Tables.

363

364 Web crawlers (e.g., written in Python) that extract web-based data were found not
365 precise enough to do this work. Therefore it was elected to undertake this manually by
366 a trained researcher who is familiar with the field, methods and data types.

367
368 Thus far, we have not retrieved data from other databases. As the database grows
369 (during the above-mentioned dedicated post-doc project), there will be an increased
370 opportunity to do dataset by dataset comparison. In the long-term we would to link to
371 other databases (e.g., Pangaea) to promote facile data access/exchange, but this is
372 beyond the current scope of the project.

373

374 **Q2. What search engines and search terms (including any refinements) were used, and**
375 **how were quality controls implemented?**

376

377 Search engine and search terms: The Google Scholar search engine was used within the
378 ETH network, which allows access to nearly all journals. The search terms used were:
379 “organic carbon in surficial surface sediments”, “TOC in surficial surface sediments”
380 and “radiocarbon/¹⁴C in surficial surface sediments”.

381

382 Quality controls: only peer-reviewed papers were used, and coordinate systems were
383 converted where necessary to the now-widely accepted WSG84 coordinate system. The
384 researcher was supported by Eglinton by screening datasets and looking for obvious
385 outliers. Additionally, an automated python script checked for outlying values to
386 provide a last external quality check by the lead author (Van der Voort).

387

388 **Q3 How many papers did the initial search yield, and how was the final subset arrived**
389 **at?**

390

391 The initial search yielded several hundred papers (>300 papers). In our reply to
392 Question 1B (Q1B, How were papers selected? Line 340) we have described how we
393 arrived at the final subset.

394

395 **Q4 When was the database accessed?**

396 The most recent update of the MOSAIC website was done this January. The website
397 always includes it's unique DOI and a timestamp of the most recent update. Users can
398 refer to this when they use the dataset. Additionally, of course, the original paper DOI
399 is provided for every single datapoint. We expect to update parameter space on a quasi-
400 annual basis and add datasets in higher frequency, which will be enabled by a dedicated
401 post-doc project starting March 1st.

402

403 **Q5 Does this database contain data from other databases?**

404 At the time of data collection, the data was acquired from the papers directly.

405

406 **Q6 What downstream processing of the data, or meta-data, was necessary? e.g. were units
407 converted, how was lat and long derived/converted to the same projection, how was a
408 position assigned to biogeographical zones etc?**

409

410 Yes, we did downstream processing, and this has now been explained more thoroughly
411 in the text. All data was converted to the standard units (e.g., Total Organic Carbon in
412 weight percent, ¹³C in permille and latitude and longitude in the WSG84 coordinate
413 system). Whenever the sampling year was detailed in the text, the Fm, percent modern
414 or 14C age could be converted to Delta ¹⁴C or vice versa. If data was given in percent
415 modern carbon or 14C age, they were also converted. The assignment to
416 biogeographical zones was done manually by the collecting researcher. For a future
417 iteration of MOSAIC, we are working on the automatic allocation of zones (both
418 biogeographical and EEZs) using Python, but this remains a work in progress.

419

420 Line 177-180 - this is admirable and will be beneficial, but at present does not exist. This
421 aspiration should be omitted from the current description. Instead, the authors should add in
422 the Data Accessibility section that updates will take place (how often? when?) and how to
423 access the latest version of the database. I assume that each iteration will have a documented
424 history and version number thats traceable? If not, this needs to be implemented from the
425 outset.

426

427 Thank you for this suggestion. We have clarified the text. On the website, the main and most
428 abundant data can be easily accessed. Using SQL, highly detailed information (e.g., cruise

429 name, sample name or name of ¹⁴C lab) can be accessed. Due to space limitations, we cannot
430 collapse a database with nearly a dozen tables into a user-friendly Web portal with a table.
431 Hence, we also provide the detailed data in SQL format, and have also provided an introduction
432 with examples on how to access specific details. We aim to update the database approximately
433 yearly, and the version and DOI is included on the website under the “How to use this app &
434 app version” tab. We have clarified this in the main text.

435

436 e 186 - can each individual datapoint be traced back to the individual source (paper)? It will be
437 important that users of the data can look at the context of each datapoint by going back to the
438 original source if necessary. in other words, is there a unique identifier that matches the data
439 value to the specific paper from which it was extracted? This is essential and needs to be
440 included if not already done so.

441

442 Yes, absolutely. We have clarified this in the text. All datapoints are accompanied by the DOI
443 (last column of the table). This way, indeed, a user can go back to the original source when
444 they're interested with just a few clicks.

445

446 Line 221 - how are submitted data quality checked? make this clear here.

447 Thank you for this comment, this is an important point and is detailed in the text. Briefly, an
448 automated preliminary check is done in Python to flag suspicious values, followed by a hands-
449 on check by an in-house expert. We have also included these details explicitly in the SI.

450

451 In more detail:

452 This script auto-checks the values of key parameters such as as latitude, longitude, carbon and
453 nitrogen content, ¹³C, ¹⁴C, CaCO₃ content, SiO₂ content and sediment texture-related
454 parameters. The auto-check produces a log file with flags for unexpected values. In turn, the
455 flags point to the exact line containing possible out-of-bound values. For example, for TOC
456 (%), if values are negative, there will be a prompt “*cannot be negative, please check*”, when
457 values are > 2 and <20 there is a prompt “*is quite high. Are you sure it is correct?*” and lastly
458 if values are > 20 there is the prompt “*value is high. Please check units*”. Each flag is
459 accompanied by a line number to locate the possibly erroneous data. These flags then trigger a
460 manual quality check of the data by an expert in-house user. We have now included all checks
461 in detail in the SI.

462

463 In other words, the work is automated to a large degree, but is overseen by an expert in-house
464 member of the ETH Biogeoscience group.

465

466 8. Line 228 - how exactly are unexpected values determined? How is this reconciled with
467 unexpected, or outlier variables, that are nevertheless real? Need to reassure the reader that the
468 data is not being sanitised to some pre-determined criteria or parameters.

469

470 Thank you for this question. With the automated Python-powered check, suspicious values are
471 only flagged (with data line number, so it's easy to locate the data), not removed or deleted.
472 This then allows for an in-house expert user to manually check the flagged data, as our
473 experience is that – while laborious – scripts do not substitute for “human” checks for
474 oceanographic consistency in the data. For instance, if the TOC values are high (> 2 and <20),
475 and there is a prompt: “*is quite high. Are you sure it is correct?*” the in-house expert will have
476 a closer look. If the values are from a zone of hypoxia or anoxia, or high biological productivity,
477 they will understand that the data is likely correct. A common issue, for example, is that TOC
478 data is provided in mg/g instead of percentages.

479 Many parameters (e.g., texture parameters % clay, silt, sand), values cannot be negative or over
480 a hundred percent, so such values are also flagged. Isotopic values can of course be negative,
481 but should fall within reasonable ranges.

482 Initially, details of this were in the python script, but to make them more accessible they have
483 now been added in the SI.

484

485 9. Data quality control - this section needs expanding, as stated earlier, to include quality
486 controls at the point of data collection. The current section only lists quality control post
487 collection. In addition, this section would benefit from some explanation/justification of the
488 detail, supported by citations where necessary/appropriate

489

490 Thank you for this comment, we have expanded and have included the details of the quality
491 control at the data collection.

492

493 10. Section 2.3.5 - it would be beneficial for the supplementary material to include an "idiots
494 guide" for how to complete a search and extract the data for a simple and more complex query

495 example. For example, what are the step through processes to extract a global dataset versus
496 just one region, or whatever is likely to be a common query. This should be made readable and
497 accessible to users that have never used SQL or programming, or that have little or no
498 experience of extracting data. The video is a useful addition in this regard, but a manual type
499 addition to the supplementary material would be helpful.

500

501

502 [Thank you for this comment, we have followed up on your suggestion and included a step-by-](#)
503 [step introductory guide for new users on how to access the MOSAIC SQL database in the SI.](#)

504

505 11. Section 3.1 - much of this section is unnecessary and not particularly helpful. the
506 description of the distribution of data is only relevant to the database as it now stands, but as
507 highlighted in the papers, the database will be updated. hence, such statements will be
508 misleading at the point of the first update. Instead, purely descriptive statistics that relate to the
509 database structure (i.e. not interpretative information) should be presented, such as the number
510 of observations for each variable, categorised by region, water depth and other column
511 headings in the database. Presently, it is hard for the reader to understand what the database
512 contains without entering the database itself. As made above (point #1), this section morphs
513 from being a database description to a paper that's interpreting the data. In my opinion, as
514 interesting as the summaries are, the latter has no place here. If the authors wish to interpret
515 the data, they should write a separate contribution and publish elsewhere

516

517 [Thank you for this comment. Briefly, you stress that this paper should be descriptive, not](#)
518 [interpretative and that sufficient details w.r.t. data points should be provided. Regarding the](#)
519 [point of descriptive vs. interpretative, we follow the line of other ESSD database papers where](#)
520 [it is common practice to show illustrative examples of data that can be extracted. We stress in](#)
521 [the main text that the examples are illustrative and not definitive, and to highlight the sorts of](#)
522 [information that are already emerging from this database. Regarding the point that sufficient](#)
523 [details w.r.t to datapoints should be provided, we included in Table 1 an overview of main](#)
524 [variables and their abundance, and in Figures 2 and 3 an overview of location and distribution](#)
525 [\(average, mean, median and n\) of all key variables.](#)

526

527 12. Section 3.2 - this can be condensed significantly, many of these points have been made in
528 the Abstract and Introduction. The text would also benefit from reaching out to other fields,
529 perhaps offering other areas that these data may be relevant to that have not received attention
530 previously.

531

532 Thank you for these comments. Thanks to your review, it has come to our attention that also
533 MOSAIC may also have relevance for the field of seafloor ecology. We have now included
534 some text on this point as well and citations to the papers by e.g. Isla and DeMaster, (2018),
535 Snelgrove et al., (2018), and of course Middelburg, (2018). Furthermore, as suggested, we have
536 also condensed this section.

537

538 13. Section 3.3 - this section is quite weak and not very compelling. It is not entirely clear
539 whether (i) the data contained in this database is a subset of the other databases mentioned, (ii)
540 how these data differ from other inventories and what the pros and cons of these data are in
541 relation to specific areas of research (maybe include reference to other databases that may form
542 good companions to these data), (iii) and why a user should opt for using these data? Some
543 aspects of these matters are listed, but only in very general terms that lack specifics. Much
544 more explicit arguments need to be made here.

545

546 Thank you for your comment. We have addressed the comments as detailed below:

547 - **Q13- I the data contained in this database is a subset of the other databases**
548 **mentioned**

549 Thank you, as per your suggestion we have expanded details on data collections so this
550 has been clarified. MOSAIC has been created by collecting the data from > 200 paper
551 publications, numerous from which data could only be extracted manually from PDFs.

552

553 - **Q13- II how these data differ from other inventories and what the pros and cons**
554 **of these data are in relation to specific areas of research (maybe include reference**
555 **to other databases that may form good companions to these data)**

556 We have explained how MOSAIC differs from other initiatives and have also included
557 the references to other databases.

558 In the section 3.3 we describe that MOSAIC differs from these and other initiatives in
559 its targeted approach with a primary focus on (i) pro-actively collating data pertinent to

560 OC burial on continental margins, (ii) upper sediment layers (nominally $< \sim 1\text{m}$) that
561 encompass early diagenetic processes and recent deposition (as opposed to down-core
562 studies that seek to reconstruct past ocean and climate conditions), and (iii) radiocarbon
563 information that bridges to equivalent databases for other carbon cycle compartments.
564 In this way, we envision that it will serve as a resource to enable “on-stop shopping”
565 for biogeochemical and sedimentological information on continental margin surficial
566 sediments. While thus far data ingested into MOSAIC has been retrieved from the
567 primary research literature, future efforts will focus on harmonizing and linking with
568 other databases in order to improve overall connectivity of information.

569

570 - **Q13 – III (iii) and why a user should opt for using these data? Some aspects of**
571 **these matters are listed, but only in very general terms that lack specifics. Much**
572 **more explicit arguments need to be made here.**

573 - Thank you for this comment, we have added specifics in section 3.3. We provide a user-
574 friendly interface which is very transparent, where sample location and data source are
575 directly provided (all DOIs are provided). MOSAIC constitutes that largest collection
576 of ocean shelf sedimentary data in this format thusfar.

577 - We anticipate that MOSAIC will serve as a key research and teaching resource for
578 biogeochemists focusing on contemporary biogeochemical processes as well as seeking
579 to interrogate sedimentary archives to develop records of past oceanographic
580 conditions.

581

582 14. Section 4 - add a sentence that states what version of the database this paper is referring
583 to/describing, and how often users can expect updates to the database (e.g. periodically,
584 annually?). I suggest it will also be advantageous to state how errors can be reported.

585

586 Thank you, we have now included this. Both new data and errors or bugs can be
587 reported to mosaic@erdw.ethz.ch. We expect to do a semi-annual update, enabled by a
588 fully funded post-doc in the Eglinton group dedicated to MOSAIC starting March 2021.

589

590 Section 5 - this section is repetitive of the sections above and does not add anything new. This
591 section needs revising to pick up from where the Introduction left off.

592

593 Thank you. We have shortened the section and revised it to pick up where the
594 Introduction left off.

595

596 16. Table 1- the database contains 8706 entries with latitude and longitude, but only about half
597 of these have a water depth associated with them - could those that do not have a water depth
598 be estimated using, for example, Google earth based on the lat and long co-ordinates? I note
599 the comment re GEBCO, but the same comment made earlier about the state of the database at
600 the point of publication versus aspirations stands.

601

602 This is correct. We have looked into connecting to the GEBCO database to also see if we can
603 auto-query the depth using the lat and lon and include this, but this is technically not trivial to
604 do. In the present iteration, the depth is not available for all but nonetheless numerous
605 datapoints, as clearly stated. Users can, of course, look up the depth information with GEBCO
606 using the lat and lon they get from their data.

607

608 Overall, I am supportive of the communication, but as the manuscript now stands it does not
609 include sufficient detail about how the data were derived and forms an incompatible mix of
610 existing versus aspirational database properties. I would see both of these as moderate
611 revisions.

612

613 Thank you for your support. We have included additional details w.r.t. how the data was
614 derived ad we have clarified the contents of the current iteration, and potential for growth (the
615 ocean is vastly expansive and complex!).

616 Referee #3 Paula Reimer

617

618 The MOSAIC database will be an extremely valuable tool for research in marine carbon
619 storage. It is easily used and expandable. I am not an expert in marine sediments but I was glad
620 to see the authors are thinking about adding on ^{14}C of sediment fraction from density
621 separations or thermal decomposition which can contribute important information on carbon
622 source which of course would require fields for the density or temperatures used for each ^{14}C
623 measurement. I found the examples useful though obviously more data is needed. I definitely
624 support publication of the MOSAIC article in ESSD after minor revisions.

625

626 Dear Prof. Reimer, Paula,
627 Thank you very much for your encouragement and positive feedback! It is very much
628 appreciated. We have incorporated all suggested revisions.

629

630 I have only a few very minor comments/corrections:

631 Line 73: Why reference Reimer et al. 2009 when the 14C archives have been updated in Reimer
632 et al 2020 (doi: 10.1017/RDC.2020.41)? Thank you - improved

633

634 Line 159-161: ‘Sediment depth profile data primarily used to examine diagenetic profiles, and
635 to constrain sedimentation rates, mixed layer depths, redox gradients, as well as to determine
636 carbon fluxes and inventories’. This is an incomplete sentence. Thank you – indeed it was, this
637 has been corrected.

638

639 Line 176-177: ‘and the stable carbon isotopic composition ($\delta^{13}\text{C}$ and 14C values) of OC’.
640 Delete ‘stable’ since 14C is not stable. Thank you – corrected.

641

642 Line 199: Presumably $\Delta^{14}\text{C}$ is not age corrected but, in any case, the date of the 14C
643 measurement would be needed to convert to Fm. If it is age corrected then the year of collection
644 is also needed.

645

646 Thank you, indeed, and whenever this possible this has been done. We have clarified this in
647 the text.

648

649

650 Line 327: ‘coupled the application. . .’ Presumably this should be ‘coupled with’ Thank you –
651 corrected.

652

653 Line 345: ‘14C’ should have a superscript 14C Thank you – corrected.

654

655

656

657 **References**

658

659 Arndt, S., Jørgensen, B. B., LaRowe, D. E., Middelburg, J. J., Pancost, R. D. and Regnier, P.:
660 Quantifying the degradation of organic matter in marine sediments: A review and synthesis,
661 *Earth-Science Rev.*, 123, 53–86, doi:10.1016/j.earscirev.2013.02.008, 2013.

662 Avelar, S., van der Voort, T. S. and Eglinton, T. I.: Relevance of carbon stocks of marine
663 sediments for national greenhouse gas inventories of maritime nations, *Carbon Balance*
664 *Manag.*, 12(1), 10, doi:10.1186/s13021-017-0077-x, 2017.

665 Bianchi, T. S.: The role of terrestrially derived organic carbon in the coastal ocean: A changing
666 paradigm and the priming effect, *Proc. Natl. Acad. Sci.*, 108(49), 19473–19481,
667 doi:10.1073/pnas.1017982108, 2011.

668 Griffith, D. R., Martin, W. R. and Eglinton, T. I.: The radiocarbon age of organic carbon in
669 marine surface sediments, *Geochim. Cosmochim. Acta*, 74(23), 6788–6800 [online] Available
670 from: <http://linkinghub.elsevier.com/retrieve/pii/S001670371000493X> (Accessed 2 August
671 2013), 2010.

672 Isla, E. and DeMaster, D. J.: Labile organic carbon dynamics in continental shelf sediments
673 after the recent collapse of the Larsen ice shelves off the eastern Antarctic Peninsula: A
674 radiochemical approach, *Geochim. Cosmochim. Acta*, 242, 34–50,
675 doi:10.1016/j.gca.2018.08.011, 2018.

676 Middelburg, J. J.: Reviews and syntheses: to the bottom of carbon processing at the seafloor,
677 *Biogeosciences*, 15(2), 413–427, doi:10.5194/bg-15-413-2018, 2018.

678 Snelgrove, P. V. R., Soetaert, K., Solan, M., Thrush, S., Wei, C. L., Danovaro, R., Fulweiler,
679 R. W., Kitazato, H., Ingole, B., Norkko, A., Parkes, R. J. and Volkenborn, N.: Global Carbon
680 Cycling on a Heterogeneous Seafloor, *Trends Ecol. Evol.*, 33(2), 96–105,
681 doi:10.1016/j.tree.2017.11.004, 2018.

682 Voort, T. S. Van Der, Mannu, U. and Blattmann, T. M.: Deconvolving the fate of carbon in
683 coastal sediments, *Geophys. Res. Lett.*, 45(June), 4134–4142, doi:10.1029/2018GL077009,
684 2018.

685