

1 Initial paper submitted to OS 25/09 2024

2

3 We thank the two reviewers for their constructive comments, which we find rather
4 encouraging to our ongoing efforts to set up SCOR working group 171 (MASIS) devoted to
5 water isotopes in sea water and to the dissolved carbon isotopic composition. The tasks of this
6 working group will include the investigation of methods to identify offsets between data sets
7 that are not related to geophysical variability, as well as to favor direct intercomparison of
8 samples and of collection, storage, measurement and validation methods. We hope that this
9 will contribute to improve the overall quality and usability of these wonderful measurements.

10 The present note had the humble goal to illustrate in two instances the magnitude of offsets
11 between available surface data sets. We chose data sets that have some overlap and that were
12 recently published. Surface data are a significant part of the available water isotopes data in
13 sea water, and present their own challenges (natural spatio-temporal variability in relation to
14 surface salinity, for example). The deep subsurface data also present very distinctive issues
15 and diagnosing the offsets between them is a challenge that is tackled by at least two groups
16 of researchers that will contribute to MASIS.

17 We will provide below detailed responses to the two reviews (preceded by Au: and in red).

18 RC1

19 General comments: This short note presents some seawater oxygen and hydrogen isotope data
20 generated in the lead author's laboratory, and conducts a comparison with other datasets from
21 the same ocean regions generated in other labs. The major finding appears to be that there are
22 troubling differences in the data, possibly instrumental or procedural in origin. Whilst
23 concerning, this is not an especially surprising finding – I think that most scientists who work
24 with such data are quite attuned to the possibility of inter-laboratory offsets, and in many
25 cases will conduct their own ad-hoc comparisons to try and make datasets consistent where
26 possible.

27 Different instrumentation and protocols could be partly to blame, in some cases, and the paper
28 outlines the possible origins of some of these differences. I suspect that long-term
29 maintenance of standards is also an issue – while all samples are supposed to be measured
30 relative to known standards (e.g. VSMOW), the cost and availability of VSMOW is such that
31 samples are virtually never measured against it directly, but instead against intermediate
32 (secondary) laboratory standards, which are themselves measured relative to e.g. VSMOW.
33 Any drift or inaccuracy in the known composition of these secondary standards will thus feed
34 through into the data.

35 **Au:** we fully agree about the issue of drifting or uncertain secondary standards, or opening
36 too many times the small VSMOW standard bottle (we had mentioned ‘uncertainties in
37 reference materials used’, but will make that clearer). This could result in biases in estimated
38 isotopic values (often too low values). There are published recommendations (from IAEA) on
39 how to keep the secondary standards for a long time free of noticeable drifts, which some
40 groups (but not all) follow. Even though, we are aware of the importance to check at relative
41 intervals whether the standards are kept in good condition, and the possibility of small errors
42 due to that still sneaking in (in $\delta^{18}\text{O}$, possibly on the order of 0.01 to 0.02 ‰).

43 The current paper does a reasonable job of highlighting these sorts of issues in the context of
44 the datasets examined, but overall the treatment is relatively superficial. The analysis
45 essentially compares a couple of datasets and considers whether differences might be “real”
46 (i.e. deriving from oceanic temporal or spatial variability), before concluding that they are
47 most likely not. One could argue that an even more useful analysis would assess all available
48 public isotope datasets and conduct some sort of crossover analysis that would tabulate
49 offsets. (I believe such an activity is being conducted at AWI Bremerhaven, and I note the
50 acknowledgement to one of the key researchers there – perhaps there is scope to ramp up that
51 dialogue and deepen the analysis presented here, especially given AWI representation
52 amongst the authorship team for this paper?). This would enable the full extent of the problem
53 to be at least quantified, and possibly its root causes to be better identified, if the derived
54 offsets were categorised by method, protocol, instrumentation etc. I realise this is a much
55 bigger job than the authors intended to undertake, but I feel it would be significantly more
56 useful.

57 **Au:** We thank the reviewer for the feedback and insight. We are aware of two such efforts
58 (one at AWI (Bremerhaven, Ge), and one at LOCEAN-LSCE (Paris, Saclay, Fr)) to
59 systematically compare a large ensemble of different (water column) water isotopes data sets,
60 and to characterize relative biases. This is a very valuable effort that we did not want to
61 duplicate in this study. We would like to point out that there is a large amount of data that are
62 surface (or near-surface/upper ocean only), in particular in tropical oceans (but also subpolar
63 gyres, and even in the Southern Ocean), and that are not included in these other investigations
64 (based on the surveys we have, probably more than half the data). The interest of our present
65 study, as we fully agree, just aimed at illustrating the issue on a few cases. Consequently, we
66 examine a few different data sets containing data rather close in time and space, albeit not as
67 close as wished for, which allows to quite well identify the systematic differences. This is
68 done to raise awareness, which does not seem to be shared by all data producers, to the best of
69 our knowledge, and to promote further crossover investigation in the recently approved
70 SCOR WG 171 – MASIS, an effort similar to what you propose.

71 Concerning ways forward, I think a minimum requirement of the paper should be a clear
72 statement on how the issue should be addressed in future. The paper alludes to some possible
73 methods that might be used to alleviate/address the problem (e.g. exchange of samples
74 between labs), but what is needed are firm recommendations and suggestions for who can
75 follow them up and (critically) who should oversee the process. For example, it is worth
76 noting that GO-SHIP has established protocols to deal with exactly this issue for other
77 variables to ensure their intercomparability, and might be well-placed to transfer/apply those
78 protocols to seawater isotope data also. Alternatively, possibly IAPSO has a role?

79 **Au:** we fully agree with that. This is one of the objectives of SCOR WG 171 – MASIS
80 (Towards best practices for Measuring and Archiving Stable Isotopes in S seawater) that two of

81 the authors co-chair, and which is just being established this month. We have already set-up
82 different data exchange protocols, and started exchanging samples (and will include the AWI-
83 Bremerhaven austral ocean effort, piloted by Alexander Haumann, who is also a member of
84 the working group). As this is now official, we have added a paragraph on it to the note.

85 Comment (not included in the note): We have also been in touch with GO-SHIP co-
86 chairpersons, and will make a more official presentation on what we plan (the aim is to move
87 ‘water isotopes’ from a parameter of class 3 to a parameter of class 2 in GO-SHIP). We are
88 also in touch with IAPSO (but for $\delta^{13}\text{C-DIC}$), and will liaise with the IOC UNESCO Ocean
89 Best Practices Group. We contemplated submitting last year a proposal on ‘water isotopes’ to
90 IAPSO, but timing was short, and we thought that it would weaken the SCOR WG proposal
91 and delay actions in too many different entities.

92 What is definitely not wanted is each lab or user conducting their own
93 intercomparison/correction exercises, since the resulting datasets (while internally more
94 consistent than before) will still not be comparable across them all, if different methods to
95 intercompare/correct are used.

96 **Au:** We fully agree. This was one of the main reasons why we decided to set up the SCOR
97 working group MASIS.

98 Overall, the paper highlights an issue that is concerning but not surprising. I have no objection
99 to the paper being published – I believe what it says is true, and the topic is an important one
100 – but a fuller treatment of the issue would be even more beneficial to the community,
101 including very specific recommendations on how it should be addressed.

102

103 **Au:** Thank you for your comment.

104

105 Minor points.

106 Title: perhaps should mention “... and suggested ways forward”, or suchlike? Highlighting
107 the issue is important, but even more useful would be outlining what needs to be done to
108 resolve it.

109 **Au:** This is an interesting suggestion, but we feel that this is outlined in some way in the
110 second part of the title, and thus plan to keep the title as is.

111 Abstract: Is written from the context of comparing LOCEAN datasets to others, which is
112 sensible (I’m sure it is what the authors’ starting point was) – but perhaps just saying
113 “intercomparing available public datasets” would be more balanced?

114 **Au:** We agree and changed the text accordingly.

115 Line 22. “carried out”?

116 **Au:** Done

117 Line 24. Punctuation is important here: "... between data sets, which exceed 0.1 in d18O and
118 0.5 in d2H, ..." – the commas matter!

119 **Au:** Done

120 Page 1 and 2. Just a stylistic thing, but these paragraphs are really long... it would help the
121 reader to break them into smaller chunks.

122 **Au:** Thank you. This is implemented.

123 Line 123. When examining offsets, it's a bit unsatisfying that the isotope data from Polarstern
124 were not collected from the same waterline as the TSG. Some quantification of the impact of
125 this would be useful, especially if the sensors/intake were at different depths and/or different
126 positions on the hull.

127 **Au:** Yes, this is unfortunate, but the isotopic samples were probably collected there, because
128 of easier access to the water line. Author GR has been involved in studies of vertical near
129 surface ocean salinity stratification in a working group set up by spatial agencies (EA and
130 NASA). In the RV Polarstern data set, we found a few instances near Svalbard (Arctic Ocean)
131 and in the Southern Ocean where there was obvious salinity and water mass origin
132 stratification. This is also one reason why we restricted the comparisons to regions where
133 such large stratification is usually not taking place. Nonetheless, a salinity stratification of (at
134 most) a few 0.01 pss is always possible between the two levels, even in the best
135 circumstances. Such instances (mostly in the tropics) were commented in Jacqueline Boutin,
136 Yi Chao, William E. Asher, Thierry Delcroix, D. Drucker, et al., Satellite and In Situ Salinity:
137 Understanding Near-Surface Stratification and Subfootprint Variability. Bulletin of the
138 American Meteorological Society, 2016, 97 (8), pp.1391-1407. 10.1175/BAMS-D-15-
139 00032.1. hal-01360859 (and references here-in, in particular, Henocq et al. (2010): Vertical
140 variability of near-surface salinity in the tropics: Consequences for L-band radiometer
141 calibration and validation. *J. Atmos. Oceanic Technol.*, **27**, 192–209,
142 doi:10.1175/2009JTECHO670.1. (in that paper, one of the authors, GR, had compared the
143 salinities from the Polarstern two water lines, which were often available at the time).

144 Line 202. ACC fronts are usually capitalised – "Polar Front" etc.

145 **Au:** Thank you. Done.

146 Various places. "pss" seems to have crept in as a salinity unit. If the data are indeed measured
147 and presented on the practical salinity scale (as stated), then the salinities are ratios and hence
148 do not have units.

149 **Au:** We explicitly mention that practical salinity has no unit. We removed other later
150 mentions in the draft, but left it on the figures, as to not leave any doubt that what we plot is
151 practical salinity, according to the 1978 'scale'.

152

153

154

155

156 • **Citation:** <https://doi.org/10.5194/egusphere-2024-3009-RC1>157 • **RC2:** '[Comment on egusphere-2024-3009](#)', Anonymous Referee #2, 05 Nov 2024 [reply](#)

158 Review

159

160 **Technical note: Large offsets between different datasets of sea water isotopic**
161 **composition: an illustration of the need to reinforce intercalibration efforts**

162 Gilles Reverdin, Claire Waelbroeck, Antje H. L. Voelker, Hanno Meyer

163

164 This technical note highlights the important consideration of systematic offsets between
165 seawater isotopic values measured using different instrumentation and/or in different
166 laboratories. Isotopic measurements have been largely underutilized to-date, and being able to
167 reliably compare data collected and/or analyzed by different parties will be key in developing
168 a cohesive understanding of the ocean isotopic system.

169

170 The authors highlight the need for establishing “well-accepted systematic guidelines for

171 data production and quality control”. Further, they advocate “enhancing scientific exchange
172 between the different institutes needs to be actively pursued, in order to reduce the errors
173 when merging different datasets”. I strongly agree with these main
174 conclusions/recommendations, and feel that ongoing, wide-spread cross-calibration between
175 institutes and research groups is the only way to achieve this.

176 **Au:** Thank you.

177

178 While I am in agreement with the overall intention of the paper, I think it is difficult to make
179 this point, as presented, using surface samples alone. Some deeper digging beyond the offsets
180 being 'rather systematic' would help strengthen the case.

181 **Au:** (also, in response to comment by RC1) We agree that there is a strong interest in digging
182 further in particular with water column (and deep ocean) data. We are aware of two such
183 efforts (one at AWI and one at LOCEAN-LSCE) to systematically compare a large ensemble
184 of different (water column) water isotopes data sets, and characterize relative biases. This is a
185 very valuable effort that we did not want to duplicate in this study. We would like to point out
186 that there is a large amount of data that are surface (or near-surface/upper ocean only), in
187 particular in the tropical oceans (but also subpolar gyres, and even Southern Ocean), and that
188 are not included in these other investigations (based on what we are aware of, probably more
189 than half the data). The interest of our present study, which is only aimed at illustrating the
190 issue on a few cases, is that we have rather close in time and space data of the different data

191 sets, albeit not as close as wished for, which allows to quite well identify the systematic
192 differences. This is done to raise awareness, which does not seem to be shared by all data
193 producers to the best of our knowledge, and promote further crossover investigation, such as
194 the one you propose.

195

196 There are a few main points in the text that I feel could be addressed more carefully and/or
197 given some more thought and discussion.

198

199 **Main point 1**

200 This technical note focuses on surface water samples. Surface waters are much more variable
201 seasonally and geographically than deep water masses due to
202 evaporation/precipitation/freezing/melting. While this is acknowledged within the paper, I'm
203 not totally convinced that the offsets observed between the relatively limited datasets are
204 lab/method based rather than seasonal and/or geographic differences.

205 **Au:** Cf response above. We are fully aware of issues of the imprint from air sea (and sea-
206 ice/liquid water) exchange, which is what motivates some scientists (including some of the
207 co-authors) to have collected these surface samples. Based on our experience (one of the
208 authors, GR, has published different studies on evaporation isotopic properties, as well as
209 investigation on rainfall or sea ice melting/freezing imprint on near-surface isotopic
210 composition), this is what motivates the restriction of the domains upon which we
211 intercompare the different data sets (tropics, sub-tropics and mid-latitudes to temperate
212 Southern Ocean areas). In these regions, and although one has to be quite careful, we have
213 some idea of what maximum seasonal cycle or interannual variability are in water isotope-
214 SSS coordinates averaged over regional scales. At first glance, they appear to be smaller than
215 when considering the surface variability at a fixed location, as salinity and isotopic
216 composition tend to co-vary seasonally. We are aware that we do not have the in-situ data to
217 fully test the impact of seasonal and interannual variability in the south-eastern Indian Ocean
218 or in the southern Atlantic Ocean in isotope-S space, but our guess is that this source of
219 variability is not the largest cause of the differences that we identify. This is an interesting
220 topic for further research in simulations of earth system models enabling water isotopes, but it
221 has not yet been done systematically, as far as we can tell (for example, simulations exist both
222 for MPI model (Xu et al., Geosci. Model Dev., 5, 809–818, 2012 [www.geosci-model-](http://www.geosci-model-dev.net/5/809/2012/)
223 [dev.net/5/809/2012/](http://www.geosci-model-dev.net/5/809/2012/), doi:10.5194/gmd-5-809-2012), in iCESM (Brady et al., 2019, Journal of
224 Advances in Modeling Earth Systems, 11, 2547–2566.
225 <https://doi.org/10.1029/2019MS001663>), and in the IPSL model).

226 Without many similar datasets demonstrating the extent of natural variability, or direct
227 replicate analyses performed at different labs, it's difficult to make a convincing case that the
228 reported differences are analytical offsets and not observed natural geographic/temporal
229 variation. In absence of direct replicate analysis/cross-calibrations, the exercise detailed in
230 this paper may be better performed with deep water samples, with less variable isotopic
231 compositions.

232 **Au:** We fully agree that deep water samples are best for that, but we unfortunately do not
233 have similarly located deep water samples in the three data sets compared here. Deep samples
234 collected at the same site are rarely available, which is why other paradigms/approaches are
235 needed, as is been done in the Southern Ocean by at least two research groups. Although we
236 have only limited understanding of the observed natural geographic/temporal variation, we
237 found that the surface comparison approach is a complementary useful approach.

238

239 **Main point 2**

240 I have a hard time recommending the application of a ‘correction’ offset between datasets
241 without a direct cross-calibration between the two labs, it is impossible to know if the
242 difference in values is an offset (from technique, reference material, or sample evaporation),
243 or a true difference.

244 While correcting for a calibration offset between labs could be acceptable with appropriate
245 inter-lab cross-calibration efforts, trying to ‘correct’ data where samples may have been
246 compromised involves considerable risk, and could instill a false sense of confidence in
247 intercomparison efforts.

248 **Au:** (also cf response to comment RC1) We fully agree that direct ‘cross-calibration efforts’
249 are required, and that is what we promote in the SCOR WG 171 – MASIS, which is just been
250 set up this month, and which is co-chaired by two of the co-authors of the paper. Indeed, it is
251 in some instances difficult to fully identify the reason of the offset. However, in the LOCEAN
252 data set, there were some clear instances, in particular for earlier data sets but even during one
253 recent cruise, of sample evaporation having taken place; in these instances, if evaporation was
254 low, an adjustment is proposed for cases when the suggested adjustment is less than 0.1/0.2 in
255 $\delta^{18}\text{O}/\delta^2\text{H}$ (with a flag specifying “probably good”). Furthermore, in the case of the RV
256 Polarstern data set analyzed at AWI-Potsdam, an offset in a reference material had been
257 identified, post initial data publication, which has been corrected before making the
258 comparison provided here. In other instances, reference material offsets happen that need to
259 be carefully identified to provide full confidence in the inter-comparison effort. We are,
260 however, aware that remaining offsets on the order of 0.01 to 0.02‰ in $\delta^{18}\text{O}$ due to
261 uncertainties in secondary reference materials might sneak in the data, even in the best
262 managed laboratories (at least, this is the case at LOCEAN).

263

264 More recently published material indicates that differences between analytical techniques (i.e.
265 IRMS vs CRDS) are insignificant (i.e. less than analytical precision). Reference material
266 errors can occur, and the only way to identify that for certain is cross-calibration between the
267 facilities in question. Unfortunately, facilities are often reluctant to spend the time/resources
268 on cross calibration, claiming that if all labs are (e.g.) referenced against VSMOW, then there
269 is no need (which is of course true, in theory, but not all labs operate the same way with
270 regards to calibration, replication, etc.). This is a problem that must be solved with buy-in
271 across the community, and a commitment to a longer-term vision of isotopic data (vs short-
272 term focus on a study from a single cruise, where offsets between labs/instruments are often
273 immaterial).

274 **Au:** We fully agree that this is an issue that must be solved with a long-term vision across the
275 community, and strongly support this buy-in vision. We'd like to point out that although there
276 are published indications that differences between analytical techniques (for $\delta^{18}\text{O}$ between
277 IRMS with CO_2 equilibration and CRDS (at least Picarro-based)) are less than analytical
278 uncertainties, this is not a universally recognized vision (L. Wassenaar, pers. communication,
279 or different papers on effects on the water isotope measurements of dissolved salts, including
280 one at LOCEAN by Benetti et al. (2017)). Thus, it might be by chance that they are so similar
281 (we believe that there are larger differences in the case of $\delta^2\text{H}$, or for other IRMS
282 measurement methods for $\delta^{18}\text{O}$, but this would need to be investigated further).

283

284 An offset resulting from sample compromise prior to analysis is the most difficult case.
285 Unless the offsets are very large, it is not truly possible to know which samples may have
286 been subject to evaporation during storage, or to what extent. Even if we can be confident that
287 some samples from a cruise were compromised, it cannot be assumed that each sample was
288 subject to the same amount of evaporation/fractionation offset. An attempt to correct some
289 number of collected samples from a dataset could have the unintended consequence of
290 unnecessarily offsetting samples within that set that were not actually subject to
291 compromise/evaporation. My opinion is that the best approach in this scenario is to discard
292 the data that has clearly been compromised (i.e. well outside of established natural variability)
293 and leave the rest alone.

294 **Au:** Cf response above. With CRDS measurements, we based this evaluation of sample
295 compromise (mainly, possible evaporation) on their d-excess value (which, if outside
296 acceptable ranges, is then flagged as probably bad), as well as on visual inspection of the
297 bottles and the caps. Users of the LOCEAN data base are free to discard samples that have
298 been lightly compromised before analysis (based on the 'probably good' quality flag). Note
299 that we have chosen to retain them in the comparisons presented in this paper, as they have no
300 statistical effect on the comparisons provided.

301 Unless a direct inter-lab cross-calibration has been performed, I would rather not see
302 'correction' offsets applied to datasets in attempt to make them more comparable. Without the
303 cross-calibration data, there is simply no way to know whether one is truly making the
304 correction they think they are.

305 **Au:** Thank you for your opinion. We promote direct inter-lab comparison calibration, but
306 unfortunately this is not always available (this was not the case for one of the data sets of this
307 paper, despite our attempts), and furthermore, an a posteriori inter-comparison would not
308 necessarily be relevant to the earlier analyzed data set. Thus, we have to conclude that it will
309 be in some instances necessary to apply 'correction' offsets, as a secondary and more
310 uncertain approach to the more direct inter-lab comparison.

311

312 --

313 **A few specific minor points:**

314

315 **L40:** Glacial melt from ice shelves can impact isotopic composition well below the surface
316 ocean – down to 800m in the Amundsen Sea, Antarctica. (Biddle et al., 2019; Hennig et al.,
317 2024; Randall-Goodwin et al., 2015)

318 **Au:** Thank you; this was added.

319

320 **L66:** More recent studies have demonstrated significantly better precision with CRDS
321 systems - on par, or better than most published IRMS data. (0.025‰ in $\delta^{18}\text{O}$. Current
322 manufacturer-stated precision for Picarro systems is $\pm 0.025\text{‰}$ in $\delta^{18}\text{O}$ and $\pm 0.1\text{‰}$ for δD).
323 Voelker et al., (2015) achieved precision of $\pm 0.06\text{‰}$ $\delta^{18}\text{O}$ (in-run precision $\pm 0.1\text{‰}$ $\delta^{18}\text{O}$;
324 $\pm 1\text{‰}$ δD), while Hennig et al., (2024) achieved precision of $\pm 0.02\text{‰}$ $\delta^{18}\text{O}$ (in-run precision
325 of $\pm 0.04\text{‰}$ $\delta^{18}\text{O}$). I'm not sure whether an advancement in technology, or a change in
326 methodology is responsible – but it doesn't seem that modern precision is meaningfully
327 different between IRMS and CRDS.

328 **Au:** Some groups in the past (still, now) provided IRMS data with similar 'high' precision,
329 but not all. I agree that there has been great progress with CRDS measurements, but there is
330 still sometimes a large issue of salt accumulation and its impact on memory effects during
331 runs. It is something that is present at LOCEAN, and for which the current solution is to run
332 reference materials through the runs with values close to the sample values analyzed. In that
333 respect, the situation is much worse for isotopic composition in salt water than in freshwater
334 (a solution that was earlier adopted at LOCEAN was to distillate the water samples prior to
335 analysis, but at great time cost). Note that memory effects could arise in IRMS measurements
336 too (but not due to salt).

337

338 **L78:** Is this significant? Other studies (Hennig et al., 2024; Walker et al., 2016) showed
339 equivalence within instrumental precision between IRMS and CRDS techniques. I'm not sure
340 one can analytically justify applying an offset to data that is smaller than analytical precision.

341 **Au:** Yes, this effect can be highly significant, unfortunately (see comment above, based on
342 experience gathered at LOCEAN on CRDS). Fortunately, this can be (partially) mitigated.

343