

1 Initial paper submitted to OS 25/09 2024

2 RC1

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

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

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

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

41 **Au:** We thank the reviewer for the feedback and insight. We are aware of two such efforts
42 (one at AWI (Bremerhaven, Ge), and one at LOCEAN-LSCE (Paris, Saclay, Fr)) to
43 systematically compare a large ensemble of different (water column) water isotopes data sets,

44 and to characterize relative biases. This is a very valuable effort that we did not want to
45 duplicate in this study. We would like to point out that there is a large amount of data that are
46 surface (or near-surface/upper ocean only), in particular in tropical oceans (but also subpolar
47 gyres, and even in the Southern Ocean), and that are not included in these other investigations
48 (based on the surveys we have, probably more than half the data). The interest of our present
49 study, as we fully agree, just aimed at illustrating the issue on a few cases. Consequently, we
50 examine a few different data sets containing data rather close in time and space, albeit not as
51 close as wished for, which allows to quite well identify the systematic differences. This is
52 done to raise awareness, which does not seem to be shared by all data producers, to the best of
53 our knowledge, and to promote further crossover investigation in the recently approved
54 SCOR WG 171 – MASIS, an effort similar to what you propose.

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

63 **Au:** we fully agree with that. This is one of the objectives of SCOR WG 171 – MASIS
64 (Towards best practices for Measuring and Archiving Stable Isotopes in S seawater) that two of
65 the authors co-chair, and which is just being established this month. We have already set-up
66 different data exchange protocols, and started exchanging samples (and will include the AWI-
67 Bremerhaven austral ocean effort, piloted by Alexander Haumann, who is also a member of
68 the working group). As this is now official, we have added a paragraph on it to the note.

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

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

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

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

86

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

88

89 Minor points.

90 Title: perhaps should mention "... and suggested ways forward", or suchlike? Highlighting
91 the issue is important, but even more useful would be outlining what needs to be done to
92 resolve it.

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

95 Abstract: Is written from the context of comparing LOCEAN datasets to others, which is
96 sensible (I'm sure it is what the authors' starting point was) – but perhaps just saying
97 "intercomparing available public datasets" would be more balanced?

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

99 Line 22. "carried out"?

100 **Au:** Done

101 Line 24. Punctuation is important here: "... between data sets, which exceed 0.1 in d18O and
102 0.5 in d2H, ..." – the commas matter!

103 **Au:** Done

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

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

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

111 **Au:** Yes, this is unfortunate, but the isotopic samples were probably collected there, because
112 of easier access to the water line. Author GR has been involved in studies of vertical near
113 surface ocean salinity stratification in a working group set up by spatial agencies (EA and
114 NASA). In the RV Polarstern data set, we found a few instances near Svalbard (Arctic Ocean)
115 and in the Southern Ocean where there was obvious salinity and water mass origin
116 stratification. This is also one reason why we restricted the comparisons to regions where
117 such large stratification is usually not taking place. Nonetheless, a salinity stratification of (at
118 most) a few 0.01 pss is always possible between the two levels, even in the best
119 circumstances. Such instances (mostly in the tropics) were commented in Jacqueline Boutin,
120 Yi Chao, William E. Asher, Thierry Delcroix, D. Drucker, et al., Satellite and In Situ Salinity:
121 Understanding Near-Surface Stratification and Subfootprint Variability. Bulletin of the
122 American Meteorological Society, 2016, 97 (8), pp.1391-1407. 10.1175/BAMS-D-15-

123 00032.1. hal-01360859 (and references here-in, in particular, Henocq et al. (2010): Vertical
124 variability of near-surface salinity in the tropics: Consequences for L-band radiometer
125 calibration and validation. *J. Atmos. Oceanic Technol.*, **27**, 192–209,
126 doi:10.1175/2009JTECHO670.1. (in that paper, one of the authors, GR, had compared the
127 salinities from the Polarstern two water lines, which were often available at the time).

128 Line 202. ACC fronts are usually capitalised – “Polar Front” etc.

129 **Au:** Thank you. Done.

130 Various places. “pss” seems to have crept in as a salinity unit. If the data are indeed measured
131 and presented on the practical salinity scale (as stated), then the salinities are ratios and hence
132 do not have units.

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

136

137

138

139

140 • **Citation:** <https://doi.org/10.5194/egusphere-2024-3009-RC1>

141 • **RC2:** ['Comment on egusphere-2024-3009'](#), Anonymous Referee #2, 05 Nov 2024 [reply](#)

142 Review

143

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

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

147

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

153

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

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

160 **Au:** Thank you.

161

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

165 **Au:** (also, in response to comment by RC1) We agree that there is a strong interest in digging
166 further in particular with water column (and deep ocean) data. We are aware of two such
167 efforts (one at AWI and one at LOCEAN-LSCE) to systematically compare a large ensemble
168 of different (water column) water isotopes data sets, and characterize relative biases. This is a
169 very valuable effort that we did not want to duplicate in this study. We would like to point out
170 that there is a large amount of data that are surface (or near-surface/upper ocean only), in
171 particular in the tropical oceans (but also subpolar gyres, and even Southern Ocean), and that
172 are not included in these other investigations (based on what we are aware of, probably more
173 than half the data). The interest of our present study, which is only aimed at illustrating the
174 issue on a few cases, is that we have rather close in time and space data of the different data
175 sets, albeit not as close as wished for, which allows to quite well identify the systematic
176 differences. This is done to raise awareness, which does not seem to be shared by all data
177 producers to the best of our knowledge, and promote further crossover investigation, such as
178 the one you propose.

179

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

182

183 **Main point 1**

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

189 **Au:** Cf response above. We are fully aware of issues of the imprint from air sea (and sea-
190 ice/liquid water) exchange, which is what motivates some scientists (including some of the
191 co-authors) to have collected these surface samples. Based on our experience (one of the
192 authors, GR, has published different studies on evaporation isotopic properties, as well as
193 investigation on rainfall or sea ice melting/freezing imprint on near-surface isotopic
194 composition), this is what motivates the restriction of the domains upon which we

195 intercompare the different data sets (tropics, sub-tropics and mid-latitudes to temperate
196 Southern Ocean areas). In these regions, and although one has to be quite careful, we have
197 some idea of what maximum seasonal cycle or interannual variability are in water isotope-
198 SSS coordinates averaged over regional scales. At first glance, they appear to be smaller than
199 when considering the surface variability at a fixed location, as salinity and isotopic
200 composition tend to co-vary seasonally. We are aware that we do not have the in-situ data to
201 fully test the impact of seasonal and interannual variability in the south-eastern Indian Ocean
202 or in the southern Atlantic Ocean in isotope-S space, but our guess is that this source of
203 variability is not the largest cause of the differences that we identify. This is an interesting
204 topic for further research in simulations of earth system models enabling water isotopes, but it
205 has not yet been done systematically, as far as we can tell (for example, simulations exist both
206 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/)
207 [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
208 Advances in Modeling Earth Systems, 11, 2547–2566.
209 <https://doi.org/10.1029/2019MS001663>), and in the IPSL model).

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

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

222

223 **Main point 2**

224 I have a hard time recommending the application of a 'correction' offset between datasets
225 without a direct cross-calibration between the two labs, it is impossible to know if the
226 difference in values is an offset (from technique, reference material, or sample evaporation),
227 or a true difference.

228 While correcting for a calibration offset between labs could be acceptable with appropriate
229 inter-lab cross-calibration efforts, trying to 'correct' data where samples may have been
230 compromised involves considerable risk, and could instill a false sense of confidence in
231 intercomparison efforts.

232 **Au:** (also cf response to comment RC1) We fully agree that direct 'cross-calibration efforts'
233 are required, and that is what we promote in the SCOR WG 171 – MASIS, which is just been
234 set up this month, and which is co-chaired by two of the co-authors of the paper. Indeed, it is
235 in some instances difficult to fully identify the reason of the offset. However, in the LOCEAN
236 data set, there were some clear instances, in particular for earlier data sets but even during one
237 recent cruise, of sample evaporation having taken place; in these instances, if evaporation was

238 low, an adjustment is proposed for cases when the suggested adjustment is less than 0.1/0.2 in
239 $\delta^{18}\text{O}/\delta^2\text{H}$ (with a flag specifying “probably good”). Furthermore, in the case of the RV
240 Polarstern data set analyzed at AWI-Potsdam, an offset in a reference material had been
241 identified, post initial data publication, which has been corrected before making the
242 comparison provided here. In other instances, reference material offsets happen that need to
243 be carefully identified to provide full confidence in the inter-comparison effort. We are,
244 however, aware that remaining offsets on the order of 0.01 to 0.02‰ in $\delta^{18}\text{O}$ due to
245 uncertainties in secondary reference materials might sneak in the data, even in the best
246 managed laboratories (at least, this is the case at LOCEAN).

247

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

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

267

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

278 **Au:** Cf response above. With CRDS measurements, we based this evaluation of sample
279 compromise (mainly, possible evaporation) on their d-excess value (which, if outside
280 acceptable ranges, is then flagged as probably bad), as well as on visual inspection of the
281 bottles and the caps. Users of the LOCEAN data base are free to discard samples that have

282 been lightly compromised before analysis (based on the ‘probably good’ quality flag). Note
283 that we have chosen to retain them in the comparisons presented in this paper, as they have no
284 statistical effect on the comparisons provided.

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

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

295

296 --

297 **A few specific minor points:**

298

299 **L40:** Glacial melt from ice shelves can impact isotopic composition well below the surface
300 ocean – down to 800m in the Amundsen Sea, Antarctica. (Biddle et al., 2019; Hennig et al.,
301 2024; Randall-Goodwin et al., 2015)

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

303

304 **L66:** More recent studies have demonstrated significantly better precision with CRDS
305 systems - on par, or better than most published IRMS data. (0.025‰ in $\delta^{18}\text{O}$. Current
306 manufacturer-stated precision for Picarro systems is $\pm 0.025\text{‰}$ in $\delta^{18}\text{O}$ and $\pm 0.1\text{‰}$ for δD).
307 Voelker et al., (2015) achieved precision of $\pm 0.06\text{‰}$ $\delta^{18}\text{O}$ (in-run precision $\pm 0.1\text{‰}$ $\delta^{18}\text{O}$;
308 $\pm 1\text{‰}$ δD), while Hennig et al., (2024) achieved precision of $\pm 0.02\text{‰}$ $\delta^{18}\text{O}$ (in-run precision
309 of $\pm 0.04\text{‰}$ $\delta^{18}\text{O}$). I’m not sure whether an advancement in technology, or a change in
310 methodology is responsible – but it doesn’t seem that modern precision is meaningfully
311 different between IRMS and CRDS.

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

321

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

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

327