Review of OS manuscript egusphere-2025-3588

Title: Metrological concepts applied to Total Alkalinity measurements in seawater: reference materials, inter-laboratory comparison and uncertainty budget

General comments

The revision has addressed most of the reviewer's concerns. A few more minor comments are mentioned below. The reviewer supports publication after they have been addressed. There is no need for another review.

Specific technical comments

(The line numbers indicate those given by the authors. They do not indicate the position exactly, but have nevertheless been kept by the reviewer.)

line	**Reviewer's comment**	**Author's comment**	**Re-review**
general	The reviewer recommends using an LLM-based AI to improve the language. In parts, the paper is difficult to read due to linguistic weaknesses, which the reviewer did not further correct.	The language has been improved in parts were it was difficult to read but without using an AI based tool.	Technically, the topic is presented properly now. I leave it to the editors to assess language.
21	Remove "potentially". Either a result is traceable to a metrological reference or it isn't. There is no in-between status.		New comment
55	The reviewer is not convinced that simply having no uncertainty budget is the sole reason why TA measurements using the conventional method are not traceable (to whatever reference). The method relies on several measured quantities the traceability of which may not be fully established or may even be inconsistent. For example, the total		New comment.

	hydrogen ion concentration is quantified through pH/potential measurements using glass electrodes. However, what is the metrological refence of those results? Primary pH buffers, the values of which include or do not include the Bates-Guggenheim convention? In fact, the pH of those buffers is defined in terms of activity, while Dickson's guide assumes the potentials are a measure of H+ ion concentration. Moreover, how are liquid junction potentials of the glass electrode considered, which also affect the measured potentials significantly? Those are difficult questions to be answered in assessing traceability of the TA measurement procedure. That said, the reviewer does not intend to question the overall paper on the basis of these traceability concerns. However, it would be expected that this point is acknowledged in the introduction and the traceability section as an open issue. In fact, it even supports the value of the proposed artificial RM.		
172	The measurement result at zero NaCl mol/kg sol is shown …	The manuscript has been changed accordingly.	Correct: The measurement result~~s~~ at zero NaCl mol kg-1 sol is shown in Fig. 1 and support**s** this reasonable assumption. Or: The measurement results at zero NaCl mol kg-1 sol ~~is~~ are shown in Fig. 1 and support this reasonable assumption.
175	Even with goodwill, Fig. 2 does not support the assumption of a linear relationship passing through	The fact that both the gravimetric and potentiometric approaches yield	"The measurements presented in Fig. 1, which

	the origin. It rather shows a square root like behavior, which is difficult to explain. Alternatively, the ΔTA values at 1, 2 and 3 mol/kg NaCl solution content indeed suggest that there is a linear relationship - which one can expect in dependence of NaCl content – but with an offset at zero NaCl content. Which raises the question, why the measured ΔTA value is zero at zero NaCl content? I suspect, that the reason for this discrepancy can be found in the different metrological references involved in the gravimetric and measured TA values. See also comments related to lines 236 and 522. Anyhow, the linear extrapolation might be used as a rough estimate for the background alkalinity. However, the authors must comment on the difficulties I have mentioned.	ΔTA=0 µmol kg−1 at zero NaCl content supports the internal consistency of the measurements. We acknowledge, however, that the data presented in Figure 2 do not perfectly support a linear relationship passing through the origin, and we agree that this aspect warrants improvement. This limitation has been clarified and further discussed in the revised manuscript (Sect. 3.1.2 and 3.3.2)	can question the linear behaviour… " I would suggest to write "Linearity of the measurement results is a rough assumption that is further discussed in section 3.3.2."
285	"This study highlighted that the determination of the homogeneity is highly dependent on the variability of the measurement method." Should rather be "This study highlighted that the robustness of the determination of the homogeneity is highly dependent on the variability of the measurement method.		New comment
288	"It was chosen to neglect the within-bottle homogeneity." Again, the authors claim compliance with ISO Guide 35 (ISO 33405, respectively); however, their homogeneity analysis	A one-way ANOVA has been performed on the results obtained from the homogeneity testing. The ANOVA results were then used to calculate the between-bottle homogeneity uncertainty based on	Add reason: "Uncertainty resulting from within-bottle inhomogeneity can usually

	appears superficial to some extent. A one-way ANOVA must be applied to account for both within-unit and between-unit homogeneity. One might decide to disregard within-unit uncertainty for the reasons mentioned by the authors. In that case, only between-bottle homogeneity should be calculated according to the (corrected) Eq. 11. Otherwise, a proper one-way ANOVA analysis is expected for the homogeneity values given in Table 4. The authors must also evaluate the repeatability standard deviation of the homogeneity with respect to the target uncertainty (see Section 7.5.1 of ISO 33405).	the corrected equation 11 (see comment below). The corresponding values of $uhom$ have been corrected accordingly in the manuscript. ISO 33405:2024 (Section 7.5.1 of the ISO), states that the repeatability standard deviation of the homogeneity study procedure should be less than one third of the target standard uncertainty of the TA measurement result for the procedure to be considered suitable. In our case, the criterion was slightly exceeded, but the results can nevertheless be regarded as a preliminary estimate of the material's homogeneity. This has been added in the discussion section 3.4.4.	be neglected for liquid refence materials." A quantitative number should be added to the discussion in 3.4.4 to support the statement that ", the (1/3) criterion was slightly exceeded", i.e. how do the measurement repeatabilities of the three batches, which can be expressed by M-within, compare to the target uncertainty.
426	Again, remove "potential". The issue of the TA-background has been appropriately discussed. It is indeed an issue that must be addressed. But it is not so significant that SI traceability of the assigned TA value must be stated as "potential". It is rather an uncertainty of the uncertainty.		New comment
450	Replace "precision" by "repeatability".		New comment
458	"Having a natural seawater reference material that is easy to collect during open-ocean oceanographic cruises …" I find it difficult to see how this proposal could be implemented in practice, or what its benefit would be. Which institution would characterize such an *in situ* RM prepared by the operator during a cruise? And if that were feasible, why would the user rely on any other RM? If the	Some oceanographic laboratories already produce home made standards, which has for interest that it can be produced in large volumes. (e.g. EuroGO-SHIP project). The artificial material could serve has a reference material for validating their measurement method before attributing a reference value. This secondary material could also be sent to reference laboratories (e.g. NMIs) for characterization. This has been added to the manuscript.	I see the necessity, but I am not yet convinced that the proposed concept of a kind of "practical traceability" using two RMs having different traceabilities has been developed in sufficient metrological depth. However, the issue is not fundamental in the context

	operator is capable of characterizing an RM, they could directly apply the same method to measure their samples.		of the paper (even though it is very fundamental in general). Thus, I consider it resolved.
586	"… lack stability": A good observation that appropriately addresses the scope of the paper.	The manuscript was changed accordingly.	? Why? This comment didn't request a change. I was just appreciating the observation.