Response to reviewers – RC1

OS manuscript egusphere-2025-3588

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

Auhors: Gaëlle Capitaine, Samir Alliouane, Thierry Cariou, Jonathan Fin, Paola Fisicaro, Thibaut Wagener

Dear reviewer,

We would like to greatly thank you for taking the time to review our manuscript, and giving us the opportunity to improve it thanks to your valuable comments and suggestions. We have carefully considered all your comments. You will find below how we addressed them in the revised manuscript.

The line numbers correspond to the reviewed manuscript with changes marked.

Line	Comment	Response
General comment 1	The traceability of the TA values and their uncertainties - which must be the core element of any RM, particularly if the authors, as they claim, worked in accordance with ISO 17034 – are inconsistent to some extent and are not sufficiently discussed. As the authors have noted themselves, TA measurements using the conventional method according to Dickson's Guide have no proper traceability. A detailed analysis of its traceability is currently lacking. Nevertheless, the authors use this method to characterize TA of the natural seawater TA. Moreover, even though the gravimetric approach to characterize the artificial RM is traceable to the SI, the quantification of impurities in NaCl is conducted with the common method, which introduces an inconsistency into the characterization. Additionally - as the authors themselves note - the artificial seawater differs chemically from natural seawater. This also affects the regression method used to determine TA and thus its associated measurement uncertainty. Consequently, measurements of seawater TA that are referred to the artificial RM may still contain significant biases, even though the comparison measurements show reasonably good agreement.	A section "6 Metrological traceability" has been added at the end of the discussion section: "Metrological traceability is defined as the "property of a measurement whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty" (JCGM 200:2012, 2012). The absence of an uncertainty budget associated to the measurement results and to the TA value of the reference materials currently distributed by the Scripps Institution of Oceanography prevents proper traceability of the measurement results. By developing an artificial reference material with a reference value accompanied by a complete uncertainty budget, as well as by providing an initial estimation of the uncertainty in TA measurement results, this study advances the establishment of

	Beyond these more fundamental issues, the overall concept of traceability that the authors seem to have in mind is not presented clearly. Traceability can fundamentally be established only to one metrological reference—either to the artificial or to the natural seawater RM. However, two very different RMs are introduced: the artificial RM, characterized gravimetrically, and the natural seawater RM, measured using the standard method according to Dickson's SOP 3a/b. The first is proposed to establish SI traceability while the latter is proposed to serve as an additional reference for quality control. However, what if this kind of quality control provides a deviating result, which is the RM to believe? In turn, if both RM provide compatible results, why is there a need for second kind of RM? The manuscript is somewhat vague with respect to those traceability issues. The authors are not expected to solve these difficult, general problems of traceability in this paper, which would likely be beyond the scope of the study that mainly aims to evaluate the RMs, which has sufficient value in its own. Nevertheless, they must discuss and contextualize both RMs in light of these traceability challenges and clearly define their respective limitations.	traceability. To fully establish traceability, it will be necessary to quantify the background alkalinity in the artificial reference material in a more robust manner, and to evaluate the uncertainty associated with the NLLS regression. A proposed traceability route, based on the two reference materials developed, is presented in Capitaine et al. (in preparation)." We agree this is a highly important subject to discuss. As this is not the core of this paper, we planned to publish a specific opinion paper on the proposal of an enhanced route of traceability for TA measurement results : Capitaine, G., Fisicaro, P., and Wagener, T.: Towards improved metrological traceability of seawater Total Alkalinity measurements: advancing assessment of Ocean Alkalinity Enhancement, In preparation." See also the added comment line 460.
General comment 2	The structure of the paper should be reconsidered. The typical format—theory, results, discussion—makes this paper rather confusing, since each of the three sections successively addresses several different topics. There are two RMs, two characterizations, homogeneity and stability studies, and two comparison measurements. For each, the theoretical background is first explained, then all results are presented together, and	Taking your comment into consideration we chose to modify the structure of the paper with three main sections – Development of reference materials – Inter-laboratory comparison – Uncertainty budget; each of them containing relevant materials & methods, results and discussion subsections. We did not separated the artificial RM from the natural one to avoid

	finally everything is revisited for discussion. As a result, the reader easily loses track of which parts belong together and is constantly forced to flip back and forth between sections. This is not a mandatory requirement to be addressed for publication. However, to the reviewer's opinion readability would be improved if the authors restructure the paper by addressing the artificial seawater RM first—covering its preparation, characterization, stability, and homogeneity, including the relevant calculation principles and results, and concluding with the discussion. The same should then be done for the natural seawater RM. Finally, the comparison measurements should be presented, allowing both RMs to be contrasted, and a proposal for traceability should be discussed in more detail, also from a practical perspective. Additionally, the paper is rather long, considering that it essentially evaluates RMs using well-established procedures. The authors should consider shortening it to some extent.	repetitions, mainly for the methods of homogeneity and stability studies. We have also added a small section on traceability (see comment above).
General comment 3	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.
47	How do uncertainty limits illustrate climatic variations in TA? Please rephrase for more clarity.	The value has been rephrased as "These values were chosen in order to obtain a 1% standard uncertainty in the computation of the carbonate ion amount content variable, enabling to highlight climatic variations in the monitoring of ocean acidification."

56	"Not fully traceable" is a strong statement that requires discussion. If characterized HCl was used for the titration, why is the measurement not traceable to the SI? This should be discussed — if not in the introduction, then elsewhere in the manuscript.	The lack of uncertainty budget attributed to the reference value prevent from the establishment of the traceability. The sentence has been changed accordingly.
59-61	The metrological terminology is somewhat imprecise. Comparability of results is achieved through traceability to the same metrological reference, not through uncertainty. Measurement uncertainty defines the limits within which differences between measurement results — or their equivalence — become meaningful (see also VIM: compatibility). Only deviations exceeding the measurement uncertainty can be regarded as significant. Similarly, the term "uncertainty of a measurement method" is incorrect — a method itself has no uncertainty; only a measured value has one. The authors should also verify whether the word "trueness" in line 56 expresses what they intend to say. According to the VIM, *trueness* refers to "the closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value." I am not sure, if this meant.	According to your comments, the sentence has been rephrased as "Moreover, the uncertainty budget of the measurement method results is required to check the compatibility of total alkalinity values." Moreover, "improving the trueness of the results" has been replaced by "assessing more robustly an eventual measurement bias".
63	ISO Guide 35 has been replaced by ISO 33405. A paper related to metrological science should not refer to outdated standards.	The manuscript has been changed accordingly.
Sect 2, l.91	Section 2.1: The purpose of this brief summary of Dickson's SOP 3b is not clear. Usually, reference to Dickson's Guide would be suffice, all the more, the paper is already quite long. If there is a reason for the repetition, it should mentioned. I assume the formulas are	As different methods exists for the measurement of seawater TA (open or closed cells, multi-step or single addition of acid), the brief description allows to clearly state on which method the paper focuses on. Indeed, the description of the

	mentioned because it is relevant for the uncertainty calculations in subsequent sections?	measurement model is needed for the uncertainty determination. This has been specified in the text.
172	The measurement result at zero NaCl mol/kg sol is shown …	The manuscript has been changed accordingly.
173	Figure 2 is mentioned in the main text before Figure 1. Figures should be cited in the order of their appearance.	Former Figure 2 now appears as Figure 1 due to the restructuration of the paper. The numbers of the figures have been replaced accordingly.
173	Replace “theory” with “reasonable assumption.”	The manuscript has been changed accordingly.
175 & 433	Even with goodwill, Fig. 2 does not support the assumption of a linear relationship passing through 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.	The fact that both the gravimetric and potentiometric approaches yield $\Delta TA=0$ 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)
179 & 183	The purpose of measuring practical salinity and dissolved nutrients should be stated.	Practical salinity is needed for computing total alkalinity with the NLLS regression and the monitoring of nutrients gives relevant information for stability assessment. This has been added to the text.
214	A reference to ISO 33405 would be more appropriate here.	The manuscript has been changed accordingly.

223	The authors claim to evaluate the proposed RMs in accordance with ISO 17034. If so, they must fulfill the experimental requirements for short-term stability testing. Using a single, undefined transport of the RM does not meet these requirements. Since this uncertainty contribution can presumably not be readily quantified, I recommend refraining from claiming that this value has been determined. Instead, it should be stated that the value represents a first estimate, while a proper evaluation according to ISO 17034 is still pending.	The manuscript has been changed accordingly.
228	"ISO Guide 35" — see comment on line 59.	The manuscript has been changed accordingly.
251	It is unclear whether the calculation of the bias introduced by NaCl impurities is used solely to correct the reference value or to quantify its contribution to the uncertainty. This must be explicitly stated to avoid confusion. In any case, the approach appears to lead to a circular argument regarding traceability. The authors aim to correct the bias and/or assign a corresponding uncertainty to the RM. To do so, they measure TA and subtract this value from the one obtained via gravimetric measurements. However, in order to measure the TA with proper traceability, they would need a characterized RM traceable to the same metrological reference as the artificial RM — which is not available except for the proposed one. If, as assumed, the authors used Dickson's SOP to measure TA, then the traceability of the bias/uncertainty is subject to the same limitations inherent to that SOP (as mentioned in the introduction). Thus, traceability of the assigned TA value of the artificial RM, or its uncertainty, respectively, is questionable.	The calculation of the bias introduced by NaCl impurities is intended to quantify the contribution of the background alkalinity to the total TA value of the reference material. This was made clearer in the text. The uncertainty associated with this term is incorporated into the overall uncertainty budget of the assigned reference value. As indicated in Equation (5), $TA_{background}$ is added to the gravimetrically derived value, not subtracted. Because the uncertainty of $TA_{background}$ has been explicitly quantified, its inclusion in the reference value does not compromise the reliability of the assignment. Regarding the question of traceability, we believe this issue does not apply in the way suggested. The limitation mentioned in the introduction refers to the fact that currently available reference materials are not fully traceable because they lack a rigorously assessed uncertainty budget.

	Fundamentally, the bias/uncertainty must be quantified independently of the RM it is intended to characterize and with respect to the same metrological reference.	
261-264	"Systematic uncertainty sources, such as those arising from the device, the operator, or the procedure, are cancelled here." This statement is not self-evident. Which uncertainties cancel out, and how are they correlated? The authors should explain this important aspect in more detail.	The sentence "Since the same operator, instrument, and procedure were used to establish the relationship between $\Delta($ $TA_{measured} - TA_{theoretical})$ and ν_{NaCl} , these parameters contribute to systematic uncertainty sources. They are cancelled when establishing a trend and do not contribute to the uncertainty of the observed slope." Has been added for clarity.
288-302 & 466-470	"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 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).	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 the corrected equation 11 (see comment below). The corresponding values of u_{hom} 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.
297	Equation 11 is incorrect. In the simplest approach, assuming within-unit homogeneity, $u_{\text{hom}} = M_{\text{between}}/n_0$, where $M_{\text{between}}$is the mean square of the TA results of the units and $n_0$is the number of measurements per unit (assuming they are equal for each unit). See, for example, Section 7.7.3 and	The equation has been corrected based on the ISO 33405.

	Annex B1 of ISO 33405.	
305-311	The equations are mutually inconsistent. u_{stab}cannot comply with both Eq. 12 and Eq. 13. I would recommend referring to Eq. 11 in Section 8.7.3 and Annex B3 of ISO 33405 instead.	For consisteny, the equation 13 used for unstable materials is now noted $u_{stab'}$.
513	Improve clarity: Do the authors mean that the artificial RM and the stabilized natural RM were prepared at approximately the same time, or one after the other?	They were prepared approximately at the same time. The manuscript has been changed accordingly.
521	Replace "calibrated" with "characterized."	The manuscript has been changed accordingly.
328	"… and is included in the reference values given above." The meaning is unclear. The authors should be more precise: do the TA values in Table 3 really include the TA contribution from NaCl impurities, meaning the bias has not been corrected, or do they mean it has been considered, meaning the values in Table 3 have been corrected for this bias? See also the comment on line 236.	The values are computed from equation 5, which indicates that $TA_{background}$, coming from the NaCl matrix, is a contribution to the TA value of the reference material, not a bias. Meaning the TA reference value is computed from the sum of the amount contents of Na_2CO_3 and $NaHCO_3$ (gravimetric information) and of $TA_{background}$ (potentiometric determination). This has been made clearer in the manuscript.
Table 3, l.340	Table 4: The authors should add the units more precisely, as not all quantities are given in µmol/kg.	The manuscript has been revised accordingly.
336 & 466-471	Table 4: It was mentioned that within-unit homogeneity was neglected; nevertheless, corresponding values are shown. Moreover, the results suggest that within-unit variability is even larger than between-bottle homogeneity, which seems unlikely. This supports the recommendation that measurement repeatability should be assessed in relation to the evaluation of homogeneity, stability and target uncertainty (also see comment on line 271).	The uncertainty of within-bottle homogeneity wasn't calculated but results of the standard deviations from the homogeneity study. Two subsections have been added to more clearly separate homogeneity & stability tests results from the uncertainty quantification. The discussion about measurement repeatability has been added in section 3.3.4.
346-350, Figure 2	"Its stability has been studied …" I recommend adding a figure to illustrate the stability results.	The mentioned stability study (Artificial solution – batch 2) has been added as Figure 1.

581 & 595	Results should not be excluded solely for statistical reasons, especially when only a small number of participants is involved. Have potential causes of deviation related to the measurement itself been investigated?	Yes, it was mentioned in the discussion of the ILC: "Laboratory 1 later reported that the acid injection system was not functioning properly during the ILC. A leak at the microvalve, leading to inconsistent acid delivery volumes, could explain the observed bias." It is now also mentioned in this results section.
Table 5	Table 6: Using Eq. 16, the values for s_L and s_r given in Table 6, $n = 3$, $p = 4$, and $(\mu) = 1.08$, I calculate $(\Delta) = 1.40$ μmol/kg. The authors should verify the values, or clarify which numbers were used in their calculation.	Thank you for checking, the value is indeed wrong. After checking the calculation, we also found the value of $(\Delta) = 1.40$ μmol/kg. The manuscript was changed accordingly.
Table 6	Tables 6 and 7 are nearly identical. I recommend combining them.	The manuscript was changed accordingly.
388-392 & 403	Since natural seawater consists of around 90 % NaCl, it is unlikely that the difference in composition between natural and artificial seawater could account for a tenfold discrepancy between the expected and observed differences in practical and absolute salinity. In any case, it is unclear why this matters. The authors should clarify the relationship between TA and salinity to make the relevance of this discrepancy for the study apparent.	As the minor ions in natural seawater are mostly bivalent (like Ca^{2+}, Mg^{2+} and SO_4^{2-}) whereas Na^+ and Cl^- are monovalent ions, it make sense that the practical salinity (which is defined by a conductivity ratio) of a solution composed only of NaCl would be less than one of a natural seawater, although it may not explain the entire discrepancy. The sentence was rephrased as "The composition of the artificial seawater being composed in high majority of NaCl, **may explain** the higher discrepancy between practical and absolute salinity observed". The following sentences have also been added: "For natural seawater, knowledge of salinity is required to determine the TA value through NLLS regression, which accounts for the competing acid–base equilibria present in seawater. In the artificial solution, the addition of NaCl to the solution background helps maintain an ionic strength similar to that of natural seawater aiming to mimic any potential dilution effect caused by the addition of HCl; although this effect is most likely negligible (Okamura et al., 2014)."

432	I doubt the validity of this method for the reasons already mentioned in the comment on line 236 et seq. It would be more appropriate to quantify the impurities affecting TA using independent measurement methods. One cannot use the same instrument or procedure intended to be calibrated with the RM to determine the bias of that RM — this constitutes circular reasoning. It becomes impossible to distinguish whether the bias originates from the RM itself (e.g., NaCl impurities) or from the instrument or measurement procedure. For instance, if the bias depends on the ionic strength of the RM, demonstrating that the ΔTA at zero NaCl is zero does not resolve the issue.	We thank the reviewer for this thoughtful comment. The determination of $TA_{background}$ was based on several independent solutions and repeated measurements. The fact that the solution with zero NaCl content showed no detectable bias from the measurement method provides confidence that, although not the most rigorous approach, this method offers a reasonable first estimation of $TA_{background}$. In addition, the consistency observed between measurements at salinities 35 (artificial solution) and 38 (natural seawater) suggests that any dependence on ionic strength is unlikely. Furthermore, the uncertainties associated with both the gravimetric preparation of the solutions and the potentiometric measurements were explicitly taken into account in the determination of $TA_{background}$, which should limit traceability issues. However, we agree that a better method could be found, and that the determination of $TA_{background}$ will necessitate further rigorous investigation this is no more clearly presented in discussion part 3.3.2.
449	Why only “linear”: There could be other types of dependencies.	The term “linear” has been removed.
451	“However, it does not allow for the accuracy verification of TA values obtained using the nonlinear least-squares regression method …” This statement is not incorrect, but it is misleading, as it implies that the evaluation method itself is the cause of the problem. In fact, if an RM with an assigned value is available, the evaluation method is not critical — any method-related bias can be	The sentence has been revised accordingly : “However, the different chemical composition compared to natural seawater prevents from using the nonlinear least-squares regression method, which is yet widely applied to natural seawater samples to correct the value considering the acid-base system in the solution.”

	compensated by the known value of the RM. The real issue lies in the different chemical composition of natural versus artificial seawater. This difference necessitates using an evaluation method (NLLS) that differs from the one used to assign the value to the artificial seawater RM (Gran's method).	
441	If a natural seawater RM is needed anyway to measure natural seawater, what is the benefit of using the artificial seawater RM?	Having a reference material such as the developed artificial solution, with a potential for a SI traceable reference value provided alongside a comprehensive uncertainty budget (in the opposite of the natural solution) as several advantages, that are describe in the paragraph above (l.440-450) (e.g. wide range of TA values, quantification of acid titrant amount content…)
458-462	"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 operator is capable of characterizing an RM, they could directly apply the same method to measure their samples.	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.
466-470	"… method's limited precision": Where has this been discussed? As mentioned above, this evaluation should indeed be addressed (following the guidance in ISO 33405.	This has been acknowledge in regard of ISO 33405 and is now discusses in sect. 3.3.4.

483	The purpose of the DIC measurements is not stated. I assume they were intended to demonstrate that the carbon content did not change over time. Consequently, any instability of the RM must result from sources other than carbon, such as silicates. The authors should not leave it to the reader to infer the reasons for including specific results in the investigation.	The manuscript was changed accordingly.
491	"… lack stability": A good observation that appropriately addresses the scope of the paper.	The manuscript was changed accordingly.
496	"… indicating potential secondary processes influencing alkalinity." Such as? It is indeed a peculiar finding that the TA results do not reflect the increase in silicate. Identifying secondary processes in natural seawater may be difficult because of its complex composition. However, the composition of the artificial seawater is known—except perhaps for the NaCl impurities—so, an evaluation of the discrepancy should at least be feasible for the artificial RM.	Secondary processes might be biological activity, pollution, or other ion exchange processes with the glass. This has been added to the text. It is difficult to evaluate the discrepancy on the artificial RM as no nutrients analyses were performed at the beginning of the experiment.
581	The potential failure should also be mentioned in the results section (see comment on line 458).	The manuscript has been corrected accordingly.