Review of Capitaine et al.

Overall assessment:

This study reports on the preparation and evaluation of two different types of reference materials for seawater total alkalinity measurements – (1) a novel artificial material in a NaCl medium with a well-characterized composition traceable to the SI and (2) a stabilized natural seawater material. The authors assessed the homogeneity and stability of the materials and quantified the overall uncertainty through a top-down approach from an interlaboratory study with 5 participants and a bottom-up approach. The work addresses the need for a traceable reference material for total alkalinity measurements and provides a detailed uncertainty budget for the most commonly used method for total alkalinity determination in seawater. The results are informative to reference material producers and advance understanding of the quality of seawater total alkalinity measurements. Although this manuscript is worthy of publication, it requires revisions to improve presentation of statistics and other results as well as clarity of language.

General comments:

Uncertainty estimates for the natural seawater reference material

The authors do not provide uncertainty estimates for the natural seawater reference material which does not have a well-characterized reference value traceable to the SI. Although information about the trueness of the reference value is required to estimate its uncertainty, it seems that this shouldn't prevent the authors from developing a partial uncertainty budget for the open-cell titration method on natural seawater samples (similar to the uncertainty budget they presented in **Table 7**). In section 4.4, the authors discuss additional contributions that would need to be considered in an uncertainty budget for the titration of natural seawater samples. Why was this not done? Presenting an uncertainty budget for natural seawater samples would be informative to understanding the likely overall uncertainty and its most important contributions in measurements of real samples.

Expanded uncertainty:

The authors should include information about the degrees of freedom when presenting estimates of expanded uncertainties, so the level of confidence associated with the chosen coverage factor can be determined.

Tables:

The authors should carefully reconsider the organization of their tables. Some tables are cluttered with too much information and should be split into separate tables (or moved to the supplementary information), while others contain redundant information and should be consolidated. Some suggestions are offered in the detailed comments.

The formatting of the table also makes it hard to read, as the table is cut off at the end of the page. Although the formatting will be edited in the final manuscript, for the benefit of the reviewer, the authors should ensure that each table fits onto a single page.

Equations:

Many equations lack proper introduction and explanation. The authors should also carefully check the manuscript and make sure proper subscripts and superscripts are used for the various equation terms.

Detailed comments:

Lines 109-113: The description of the process is unclear. It sounds like a total of 35 liters of seawater was collected from various depths and filled into two containers. From those two containers, 25 liters of seawater were drawn and filled into a single container to produce a single batch of seawater. Was the seawater homogenized before filling into the two containers and/or again after filling into the 25 liter container?

Line 117: "Artificial seawater" is a misnomer because the background medium is sodium chloride only without the other major seawater salts.

Table 1:

Are the percent purity values listed in the table from the manufacturer or from the assay results at NMIJ and SMU? If they are assay values, they should be listed in a separate column or a separate table rather than in parentheses after the manufacturer name.

What are the likely impurities in the salts, and were the impurities assessed?

Consider changing the phrasing in the caption from "artificial solution for total alkalinity reference material" to "artificial total alkalinity reference material."

Two batches of artificial RM were produced, yet Table 1 only lists amount contents and molalities for one batch. Which batch do these values refer to? Also consider combining **Table 1** with **Table 3**. The background alkalinity of the NaCl should also be explicitly listed in the consolidated table.

Line 122: The use of pH_T may not be appropriate in a NaCl medium as it does not contain sulfate. The temperature and dissociation constants used to calculate pH should also be given. The text indicates that the pH was estimated roughly based on a Bjerrum plot. The authors should provide a more precisely calculated value (especially if listing the pH in **Table 1**) or exclude the information about pH altogether, as it isn't strictly necessary to report for this reference material.

Line 159: "Material described in Appendix B" – Is this referring to the HCl standardized at SMU? Please state explicitly to avoid confusion and reference Appendix B for the details.

Lines 163-164 and **Lines 523-524**:

Lines 523-524 indicates the possibility of background alkalinity in the other salts such as $NaHCO_3$ and Na_2CO_3. Was this assessed?

What would be the intercept in **Fig. 2** if the linear regression was not forced to zero? Might the choice to force the regression to zero discard information about the background alkalinity from the $NaHCO_3$ and Na_2CO_3?

Line 190, 193 – "Means of standard deviation" – If pooling standard deviations with uniform sample sizes, it should be calculated as the square root of the mean of the variances.

Line 201-204: Change phrasing to "ratio of the slope to the standard deviation of the slope." Also consider changing the notation so that the Student's t-value is not confused with *t* for time. α and 0 should be in subscripts. This comment also applies to **Table 4.**

Equation 8: Aren't the salts added as stock solutions? In this case, **Eq. 8** should have m_{stock} instead of m_{salt} . m_{total} should be the sum of the stock solutions plus additional water rather than the sum of the salts and water.

Line 261: "To maximize the uncertainty of the slope" suggests that the goal was to have a larger uncertainty. I think what was meant was that the first approach with the larger uncertainty estimate was selected as the more conservative estimate of the uncertainty of the slope.

Equation 11: The equation for the homogeneity uncertainty does not make sense to me. It appears to be a standard deviation of the mean, but if so, it should be $s/\sqrt{n}$. However, this would not make sense either as the between-bottle variability was estimated differently for the different batches—some batches using the standard deviation of single measurements from different bottles and another batch using the standard deviation of the bottle means from repeatability measurements. It also does not make sense why the within-bottle homogeneity was neglected in the overall homogeneity uncertainty, as the within-bottle homogeneity was explicitly estimated and listed in **Table 4.**

The observed between bottle variance should be a sum of the within bottle variance and the homogeneity variance. If the between bottle variance is calculated as the standard deviation of the means from repeatability measurements in different bottles, then

$$s_{obs,bet-btl} = \sqrt{u_{\text{hom}}^2 + \frac{s_{repeatability}^2}{n}}$$

where *n* is the number of repeatability measurements within a single bottle.

As the other reviewer noted, the within and between bottle homogeneity components can be evaluated with ANOVA according to ISO Guide 35. It would be beneficial for many readers who do not have access to the ISO documents to derive these equations at a high level.

Eq. 12 and 13: The equations for the stability uncertainty require more explanation for the reader. Two equations are used. In the case of no significant trend, the stability uncertainty only has one contribution from the uncertainty of the slope b_1, while for cases with significant trends, the stability uncertainty has an additional rectangular distribution component.

Also, what value is used for time *t*? And as noted before, this notation can be confused with the Student's *t* value.

Line 311-313: The phrasing in this sentence is confusing. The median of the set of means (from repeatability measurements made by each participant) was calculated for two different materials—natural seawater and the artificial RM (Batch 1).

Line 325: I recommend replacing "samples" with "materials" to be clear that it was two different materials being analyzed and not two bottles.

Equation 14: How is the mean $\overline{(X_l - Y_l)}$ calculated? Is it the mean difference from the 5 participants? Please clarify in the text. Why does the mean have a different subscript *l* instead of *i*, and what does it indicate?

Equation 15: There seems to be missing text that should precede this equation. The text following the equation states that s_r is the intra-laboratory standard deviation divided by the square root of the mean number of replicates – this should be explicitly written in the equation.

Equations 15 and 16: Both of these equations need proper introduction for the reader. They are based on the equations in ISO 21748. As some readers may not have access to the ISO documents, it would be helpful to provide an explanation of these equations and how the inter and intra-laboratory standard deviations are calculated.

The term $u(\hat{\mu})$ needs further explanation than simply "standard deviation of the certified reference value." It is the uncertainty of the reference value which includes contributions from the characterization of the salts, the homogeneity, and stability (**Equation 6**). The notation should be revised so that it is consistent with **Equation 6.**

Table 4: The organization of this table is very confusing. Some entries are standard uncertainties with units of μmol kg^{-1}, while others are not. The caption indicates all numerical values have units of μmol kg^{-1}, but this is not true for parameters such as the slope, slope standard deviation, and Student *t* values. The stability uncertainty (from **Equation 12** and **13**) are not listed in this table. At the very least, the authors should clearly indicate in the table which parameters are standard uncertainties and include the appropriate units for each parameter. A better approach, I think, is to limit the table to only one type of information (e.g., the standard uncertainties associated with homogeneity, stability over time, and stability to transport). Additional details on the stability evaluation (e.g., the slope, *t*-tests, etc.) can be described in a separate table. The

authors could also consider combining the information on the homogeneity and stability uncertainties from **Table 4** and **Table 5,** although the natural seawater reference material does not have a certified value and an associated uncertainty.

Table 5: Although **Equation 11** will need to be revised, I will point out that the values listed for the homogeneity uncertainty do not agree with **Equation 11** if using the between bottle standard deviations in Table 4 as s and $N = 3$. The authors should check their calculations in the tables.

Line 461: S_R is the reproducibility standard deviation. This term should be introduced and defined much earlier in Section 2.4.1, where it is used in **Equation 16.**

Line 464: Replace "precision" with "reproducibility standard deviation" to be clear what quantity is being reported. The term "precision" should be reserved for qualitative descriptions. It may also be informative to report the reproducibility standard deviation excluding Laboratory 1, as their measurements were discovered to have a systematic error due to malfunctioning of the titrant delivery on their measurement system.

Table 6: The numeric values should not be left in E+00 notation form The units also need to be specified for the standard uncertainties.

Section 3.3: This section deserves more discussion of the results rather than just a description of the data contained in **Table 6** and **Table 7.** Consider splitting this section into two—one discussing the top-down uncertainty estimates and the other discussing the bottom-up estimates.

Table 7: The formatting of this table needs much reworking to improve readability. The last column lists the individual standard uncertainties for the sub-sources of uncertainty and a combined standard uncertainty for the input parameter all in the same column. These should be separate columns. Other information is also needed such as the sensitivity coefficients used in the uncertainty propagation, and the degrees of freedom for each uncertainty contribution. Consider including a more condensed table of the uncertainty budget in the main manuscript and a more detailed version in the supplementary information.

Lines 578-593: This section discussed leaching of silicate from the borosilicate glass bottles as a potential cause for instability in some batches of reference material. It would be beneficial for other reference material producers and for future investigations to provide more details on the specifications of the borosilicate glass used (such as the manufacturer and coefficient of linear expansion of the glass), as well as any cleaning procedures performed before bottling. Were the bottles cleaned in any way? **Line 140** states that Schott borosilicate bottles were used for Batch 2 of the artificial reference material. What about the other batches?

Fig. 4: This figure is rather confusing and not very informative. Although it highlights some major sources uncertainty such as the measured potential and the volume of acid delivered, it doesn't include all the sources of uncertainty in **Table 7** and their magnitudes. A bar graph providing a visual summary of **Table 7** may be a better choice.

Equation A2:

The total hydrogen ion concentration $[H^+]_T$ on the total pH scale includes free hydrogen ions and bisulfate ions only (Dickson, 1993). $[H^+] + [HSO_4^-] + [HF]$ is the total hydrogen ion concentration on the ***seawater*** pH scale. The notation should be revised in this equation.