

l. 97 13% of the total dataset (I assume total refers to the 1981-2024, but as phrased the 13% is only for the data before 1993; whereas Fig. 3 suggests that it is the total number of ME data, that were collected in 1984-2000). Why separate this from the later period 1993-2010 (line 98, for which the 40% cited is only relative to the 1993-2010 data collection period). It would likely be wise to unify the way to report the data for each CTD type and period. Personally, I would cite ME CTD use for the total period 1984 to 2010.

Revised and corrected

On Fig. 4: I assume that the two profiles shown on each plot are for the up and down profiles respectively. They should be shown with different line sets (for example full and hyphenated), as well as different colors for T and S. Actually, this raises the question on whether the profiles stored are only the downward profiles, or are both the downward and upward profiles stored (and if so separately?)

Figure 2 was revised according to the suggestions. In BaRDO, descending CTD profiles are generally stored. However, in some cases, ascending profiles are also stored. Information indicating whether a profile corresponds to an ascent or a descent is stored in a specific field within the station table. The original data, which can be used for subsequent reprocessing (if required), remains stored in a file structure.

l. 140, add a reference to the mismatches for SBE19 data (it has been described in different papers, as well as ways to minimize it through signal processing). If I understand correctly, here the choice is made to flag the data (through the spikes, gradient and density inversion QC1), and not try to minimize the effect.

During station processing, all possible efforts were made to correct or minimize the occurrence of spikes in the SBE19 data through signal processing using different parameter settings. Nevertheless, in several cases, spikes could not be entirely removed without modifying the original profile structure. Consequently, the affected levels were flagged (e.g., through spike, gradient, and density inversion tests) instead of editing the profiles by interpolation.

l. 174: 'The structure of the database tables...'

Revised and corrected

In Fig. 6, last line of stations. I believe what is meant is TempScaleID

The last line ("TempScaleID") was checked and found to be correctly displayed.

l. 190-192: interestingly, non SBE instrument data are also processed with the standard Sea-Bird data processing. Is there some specific choices that had to be made, which are dependent on each instrument?

In most cases, standard routines such as filter, alignctd, and loopedit were applied using different parameter settings depending on the instrument and environmental conditions. For example, for the ME CTD, conductivity filters with time constants between 0.25 and 0.5 s and a pressure filter of 0.15 s were applied, together with alignctd values similar to those used for the SBE19 (e.g., -0.63 s). In the case of the NV CTD, larger values were used for the pressure filter, in many cases up to 1.0 s.

Table 2 column 'calibration date'. I am not sure that I fully understand what is reported. For example, SBE19, 1268 reported 'Jan 1993-Dec 1997". Does it mean that there were two calibration dates in jan 1993 and dec 1997. (in that case, probably title of column should be 'Calibration dates'. For ME and NBIII/V, only year (or years) reported, not a month. Does it mean that there were no known months. Also, near the end, 6214/4687 or 6218/4684, no date (means that there was no calibration, even when the sensors were provided?)

Table 2 was completed and clarified accordingly. In particular, the column title was revised to "Calibration dates". For the ME and NBIII/V instruments, information regarding the calibration months is not available, and therefore only the corresponding years are reported.

I. 207: How are 'outliers' identified here? (and are they or not included when calculating Q1 and Q3 quartiles?)

Outliers were identified using values lying more than $1.5 \times \text{IQR}$ from the first and third quartiles. Quartiles (Q1 and Q3) were calculated using the complete dataset, including the values identified as outliers.

In 3.2.1: these comparison of pairs of sensors are interesting. It would be probably even more useful to know whether some (small) systematic differences could be identified, and whether they could have been corrected. For example, on Fig. 7, one gets a sense (maybe wrong) that 4727 would tend to have positive T difference compared to other sensors. On the other hand, an information that is not provided in Fig. 7 and in the chapter, is how far in time were these comparisons of sensors done relative to when calibration is provided (of course, one can imagine that it is highly variable, but maybe it would be valuable to select in the comparisons only pairs of sensors that have not been recently been calibrated).

We agree that this is a very interesting and valuable point. However, a detailed assessment of possible systematic biases between individual sensors, including their relationship with calibration timing and potential corrections, is beyond the scope of the present dataset description paper.

3.3 I am not 100% sure that the notion of spike is clearly defined. How does one define a spike (isolated data point relative to the average of what is below or above? Or a certain thickness...). I am aware that it is in the V1.5 version of the software used, but

more explanations are needed. Otherwise, one does not understand which flag is activated, and why. Also, Figure 10 shows a huge number of flags in Temp in the spring-summer season (in particular November). What is it about? Seeing after on Figure 11, where flagged temperature points happen, I am tempted to say that I don't think that in most cases there is something 'wrong' that would need to be flagged in this season. In the example plotted it seems to be temperature inversions or spikes in the thermocline, which can happen for a whole host of natural reasons. On the other hand, if not present in SBE9plus (or on XBT profiles in the same region), this raises a question on the quality of the SBE19 profiles (maybe the pressure-dependent averaging, or temperature-lag corrections?). I also wondered whether the difference in flagged levels mentioned between SBE19 and SBE9plus is for "='Total'. It would be more instructive to have it reported separately for Temp and Sal data.

Regarding the first point, the software documentation effectively defines a spike as:

$$|V2 - (V3 + V1)/2| - |V1 - V3|/2 > V_threshold$$

that is, based on the difference between a given value and the values immediately above and below it.

On the other hand, the high percentage of flagged temperature levels during spring may indeed be related, as suggested, to density inversions produced by thermocline spikes and, clearly, to the large proportion of stations acquired with the SBE19 CTDs. In these instruments, even after applying corrections with different lag values, spikes cannot always be completely removed. Therefore, following a conservative criterion, these levels were kept flagged instead of being modified or interpolated. Ultimately, it is up to the user to decide whether or not to include these flagged data in a specific analysis.

We apologize if this was not sufficiently clear in the manuscript. An additional bibliographic reference providing a detailed description of the applied QC procedures has been included in the revised version.

The flagged levels are already reported separately for temperature, salinity, and total flags. However, the analysis was not further discriminated according to CTD model (e.g., SBE19 and SBE9plus).

I. 328: I am not sure that I understand the last part of the sentence.

"On the other hand, during cold months (June to October), when the vertical stratification substantially decreases, the percentage of levels marked is less than the total % of levels for each month."

The sentence was revised and rewritten for clarity.

Fig. 10: labels missing on vertical scale (possibly, 1, 2, 3, ...%), and is the scale the same for Sal and Temp marks?. What is meant by the black dots 'Total' that are sometimes less than the sum of Sal and Temp Marks (such as in Nov to May)

The missing labels on the vertical scale were added. The text and figure caption were revised to clarify. In Figure 10, the black dots ("Total") represent the percentage of sampled levels for each month relative to the complete dataset, and not the total number of flagged levels.

Table 4 : What is meant by Gradient in threshold (it should be in /bar or /dbar or /m?).

Gradient thresholds expressed in units per dbar. The sentence was rewritten.