

## ***Interactive comment on “Data recovery of A06 and A07 WOCE cruises” by N. M. Fajar et al.***

### **Anonymous Referee #4**

Received and published: 15 December 2011

This manuscript attempts to recover carbon data from two cruises A06 and A07 that were not included in the GLODAP database because they were deemed to be of insufficient quality based on crossover analysis methods. The authors use a 3-DwLMR multiple regression method of Velo et al. 2011 to calculate the corrections and use CT/AT ratios to confirm that the revised data are good.

I have three primary concerns with the manuscript and accompanying data. First, the grammar, spelling and overall descriptions are confusing and insufficient to fully understand what has been done. In some cases the grammar problems actually make it impossible to accurately tell what the authors did. Even without the grammar problems, the description of the MLR was not sufficient to really understand what was really done. I realize that the method is described in Velo et al. 2011, but there should be more details here since that is the justification for this whole work.

C42

My second concern is with the actual application of the MLR. The premise of the MLR is that the AT values can be calculated from the parameters used in the fit. I do not understand how this can work in this case if the original AT data are inaccurate. If I understand correctly, the authors use a moving window of AT data to determine a MLR fit, then turn around and use the parameters to calculate new AT values. Note, given the very brief description of the approach I may have missed some aspect of the approach, but that is why my first point was a better description of the technique. Assuming I understand correctly, any biases in the original data would still be in the “corrected data”. On top of that, any biases in the fit parameters would also be in the results. The authors then use the calculated AT data with the CT/AT ratio to correct the DIC data. Thus, any biases with the AT data will be translated to the CT. In my view this whole approach is flawed. It may generate smoother data, but it doesn’t necessarily generate better data.

My third concern is the “correction” of the pH data. The authors state that the pH data are “not very reliable” so they calculated the pH values using the original AT and CT data. Aside from the issue of using original AT and CT data versus their supposedly improved “corrected” data, I have a more basic problem with presenting purely calculated values in a measurement dataset. If the pH data are bad, throw them out. If people want to calculate pH they can always do that later with the AT and CT data provided. Including calculated pH data does not add any value to the dataset and gives the mistaken impression that this is a third independent parameter to check the other values. It is not independent and provides no additional information.

In summary, I do not think this manuscript and accompanying data should be accepted without a re-evaluation of the approaches.

Interactive comment on Earth Syst. Sci. Data Discuss., 4, 99, 2011.

C43