

## ***Interactive comment on “Role of regression model selection and station distribution on the estimation of oceanic anthropogenic carbon change by eMLR” by Y. Plancherel et al.***

**Anonymous Referee #1**

Received and published: 2 December 2012

Review of Plancherel, Rodgers, Key, Jacobson and Sarmiento; "Role of regression model selection and station distribution on the estimation of ocean anthropogenic carbon change by eMLR":

This manuscript by Plancherel et al. uses model data to investigate uncertainties in estimates of decadal changes in dissolved inorganic carbon with the extended multiple linear regression (eMLR) method. This represents an important piece of work, particularly since eMLR is one of very few methods to calculate the decadal uptake of carbon by the oceans due to anthropogenic CO<sub>2</sub> emissions to the atmosphere. Potential biases and uncertainties in the observationally based estimates are not quite evaluated

C6129

yet; the manuscript by Plancherel et al., tackles this issue with a very careful analysis, which is both highly relevant and significant. The manuscript is very well written and supplied with illustrative (although sometimes complex) figures. The result of the analysis is highly relevant and will be of great help to the community aiming at quantifying the decadal change in ocean DIC. The manuscript certainly deserves to be published in Biogeosciences. I recommend publication of the manuscript after attention to some comments, see below.

Major comments: The authors neglect analytical errors in their analysis; "To isolate the effect of regression model selection from other sources of error, the synthetic data are assumed free of measurement errors throughout this work". I can certainly see the advantage of doing this in order to separate out analytical errors from other errors, as discussed in the paper. The authors mention in the end of the paper that analytical uncertainties (and a few other issues) will have to be considered as well. However, the authors spend a considerable amount of text on discussing different eMLR models that best represent the "true" change in DIC. I feel that without any quantitative discussions on the role of analytical uncertainties, the value of this discussion is reduced. Obviously, analytical uncertainties varies between cruises, but for an analysis where, for instance, GLODAP data is compared to CLIVAR data the average uncertainties of the various variables is relatively well known so that this could be factored in into the analysis. On page 14610, line 16 it is stated that "Models with the lowest overall AIC values tend to be the more complex ones". Even though this is true, it would be interesting to know if this is also true when analytical uncertainties are weighted into the analysis.

Minor comments: Page 14601, line 14: "Absolute errors due to mapping are small and result in maximum vertically integrated column inventory biases of order  $\pm 10 \text{ mol m}^{-2}$ ". Agreed that  $10 \text{ mol m}^{-2}$  is a small number in relation to the column inventory of DIC. However, this is about the same amount that can be expected as the decadal uptake of DIC in the north Atlantic, so it is really a relatively large number, i.e. an error of the same magnitude as the signal you are trying to determine. This might be significant

C6130

since the station distribution is different between the “GLODAP” and “CLIVAR” stations. Page 14609, line 8: Similar to the previous point; the integrated mapping errors might be small, but on a regional scale it seems to me that they can be significant. Figure 1d: It is difficult to understand and believe that the gridding routine for the CLIVAR station network can yield such a good fit with the “true” data for some areas. This is particularly the case for the Labrador Sea where the CLIVAR station net don’t have any stations, yet the difference between the mapped and true DeltaC is relatively modest.

---

Interactive comment on Biogeosciences Discuss., 9, 14589, 2012.

C6131