**Biogeosciences
Discussions**

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.

A. Olsen (Referee)

are.olsen@gfi.uib.no

Received and published: 4 December 2012

Plancherel and co-workers provide a much needed and comprehensive model- based analysis of the performance of the eMLR approach for determination of anthropogenic CO2 on a basin scale in the North Atlantic. I recommend publication in Biogeosciences given due attention to the following comments:

page 14591, line 4. Please replace "Independant" to "Independent"

page 14591, line 19. Please change "cylcing" with "cycling"

page 14591, line 26. "Difficutlies", spelled wrong.

page 14592, line 11. Actually the eMLR type of implementation of Wallace's (1995) MLR method, wasn't first proposed by Friis et al. (2005) but by Sonnerup, R. E., P. D. Quay, and A. P. McNichol (2000), The Indian Ocean 13C Suess effect, Global Biogeochem. Cycles, 14, 903– 916. It would be appropriate to also cite this paper.

page 14592, last section states (indirectly) that the model represents adequately "natural temporal and spatial variability patterns that are consistent with many observed climate processes on a variety of time and space scales." This is actually quite critical as the many studies from the NA, (as cited in the paper and also Metzl et al., BGC 2010) have shown that natural processes introduces significant carbon cycle variations over a few years. This statement must therefore be substantiated, and it would be a natural component of this paper to show that the model reproduces at least some of this and that the eMLR approach are able to distinguish between these quite significant variations and the anthropogenic imprint.

page 14594, following eq (2), define also "T"

page 14597, line 3-4. The synthetic data are assumed free of measurement errors. As the authors should certainly be aware, this is not quite true. All measured data carry error, and some types of data carry more than others. A first-order analysis of the effect of this issue on the selection of MLR is certainly in place, in order to unravel whether the effects discussed in the paper are inferior to, similar to, or larger than the effects of measurement error and how these vary from variable to variable.

page 14598, line 16. Please provide some details on the "prescribed atmospheric carbon dioxide transient boundary condition". If I get this right, the atmospheric CO2 concentrations since the preindustrial are used as boundary, and not the emissions. This will introduce too little anthropogenic CO2 into the earth system as terrestrial and ocean uptake are already (implicitly) accounted for. Please discuss how this may affect the results that are presented.

page 14599. Regarding the Akaike Information Criterion. I am not all that familiar

with this term, nor do I expect the audience of this paper to be. Please provide some information of what is considered good and acceptable values and what is not, so as the reader may get a bit more intuitive understanding of the implications of the numbers that appear in the paper.

page 14603, line 25-26. "Continuity of the regression formulae should thus be used as a model selection criterion in addition to standard measures of statistical misfit." Would it be possible to express this "continuity criterion" more explicitly?

page 14605, line 24. "Overall, nitrate and AOU were the least often selected variables in the formulae." Please discuss, briefly, the reasons for this preference.

page 14608, discussion of effects of seasonal cycle. The seasonal cycle of the errors is not fully aligned with seasonal MLD and PP cycles. Initialization of spring bloom occurs typically in April, May, June, and from the arguments presented in the paper I would expect greater variations/errors in these months. Please explore seasonal cycle of model in order to ensure that it is well aligned with the seasonal cycle as retrieved from various data sources.

page 14612. lines 25-30. Actually, it appears from the figure that the AIC produces the best accumulated profile.

page 14613, lines 2-4. Might a more rigid procedure for selection be suggested?

page 14614, line 21, please replace "formlae" with "formulae"

page 14616, lines 8-10: "Given the available sample distribution, differences in representativeness of the sampling grids have a larger effect than changes in the seasonal distribution of the samples." This statement should be substantiated, by referring to a figure that actually shows this. From Fig 5, it appears as if the opposite is the case.

page 14617, lines 11-13. I do not completely understand the argument made here. Please consider to make this more explicit.

C6240

Finally. The authors may consider adding a third strategy to their approach. Use of regression formulae with sets of explanatory variables that are allowed to vary as a function of depth, but not in time.