2nd review of Gregor et al: Empirical methods for the estimation of Southern Ocean CO2: Support Vector and Random Forest Regression. Submitted to Biogeosciences

Response to previous concerns:

In my previous assessment, I have raised 3 main concerns. The first related to the many new terms that are not explained in the methods section, the second related to the lack of validation and assessment of the results and the third related to the unknown consequences of adding geographical proxy data.

The authors have substantially revised their methods section. It is now much easier to understand how random forest and support vector regression work. The authors did an excellent job explaining all previously undefined wording. As previously mentioned: Understanding a method means trusting a method. I have now much more trust in both methods and the results of the manuscript. Hence, I have no further reservations regarding the methods section.

The authors have further added additional validation of the methods in the revised manuscript. The authors added both the temporal and spatial error in Figure 5 as well as a sectoral error analysis in table 2. This adds additional confidence. Therefore, I have no more reservations regarding the validation of the results.

Lastly, the author also addressed my last point of concern. Namely, the discussion regarding the inclusion of coordinates as proxy variables. Despite the authors describing the changes made in their response to reviewer letter as "addressed a little better", I would even go a step further and say this is "addressed much better". I however have found the paragraph in the introduction – not section 2.3 - but that is of no concern (in fact I think the paragraph should stay where it is)

In summary, I am happy to say that the authors have adequately addressed all my previously raised concerns and substantially improved their manuscript.

Recommendation:

Based on the revision, I am happy to **recommend the manuscript for publication in Biogeosciences**. I have gathered a few technical points below that I would like the authors to consider before publication. Some of them editorial, i.e. concerning spelling etc. and some concerning the wording used.

Specific comments:

Page 1 Line 5, page 17 line 9 and page 21 line 10: I believe the authors are too negative here. Based on the evidence presented, I would not say the SOM-FFN "outperforms" the other methods. This is also along the lines of a comment made by reviewer 2 in the first round of reviews. The RMSE values are so similar and close that I would not say one method is better than the other, but rather say they "depict errors of similar magnitude".

Page 1 line 17: "Lastly, …"

Page 2 line 32: "… sensing, Mountrakis et al., 2011 found …"

Page 12 line 5: "Importantly, …"

Page 18 line 25: "(Figure 6 b,c)"

Page 21 line 4: I would certainly say they are complementary.

In general: Map plots seem to miss the last (or first) longitude entry (hence the white stripe). It often helps to plot the array 2x in longitude, i.e. from 0-720 degrees rather than 0-360 degrees. Then the stripe does not appear.