Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-46-RC1, 2019

Geoscientific Model Development Discussions

Interactive comment on "A comparative assessment of the uncertainties of global surface-ocean CO_2 estimates using a machine learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall?" *by* Luke Gregor et al.

Peter Landschützer (Referee)

peter.landschuetzer@mpimet.mpg.de

Received and published: 24 April 2019

Review of Gregor et al: A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine learning ensemble (CSIRML6 version 2019a) – have we hit the wall? Submitted to GMDD

Hamburg, April, 24th 2019, Reviewer: Peter Landschützer

Summary:

Gregor and colleagues present an impressive and comprehensive study, comparing the performance of various machine learning-based regression approaches in combination with different ocean-biome combinations. The authors compare their new estimates with the current “state-of-the-art” methods represented in the SOCOM intercomparison project and a wide range of independent and novel validation data. Using this impressive set of data, the authors ask the question, whether we have “hit the wall” in our accuracy to reconstruct the ocean carbon sink, and in what way we may further improve in the future.

Strengths:

I am particularly impressed by the amount of data and experiments used by the authors. And despite this vast amount of information, the manuscript is clearly written and easy to follow. The study provides a step forward compared to other existing intercomparison studies (e.g. the cited papers of Rödenbeck et al 2015 and Ritter et al 2017) as it compares a more consistent set (or ensemble) of estimates, all created within the same regions and with the same observational dataset (SOCATv5). Additionally, I am not aware of any other study that makes use of such a large set of independent estimates (GLODAPv2, SOCCOM, etc.) to validate their results. All of the above are significant steps forward and provide a well-suited set-up for answering the research questions posed by the authors.

Weaknesses:

I have not encountered any major weakness.

Recommendation:

This study is a significant contribution to our scientific understanding of current “state-of-the-art” observation-based pCO2 and air-sea CO2 flux estimates, their limitations in space and time as well as potential pathways for future improvement. The study will instantly be of interest and benefit ongoing carbon cycle assessment studies, such as the Regional Carbon Cycle Assessment and Processes phase 2 (RECCAP2), the Global Carbon Budget by the Global Carbon Project and last but not least, the IPCC assessment reports. I therefore recommend timely publication of manuscript. In the following, I have only listed a few minor/technical comments to be considered by the authors.

Specific and minor comments to the text:

Page 2 line 61: “did not identify” – I would rather say that SOCOM “did not look for” the best method. Unlike this study, the SOCOM comparison was slightly more difficult as some methods were still based on older observational datasets (e.g. LDEOv1) which made an objective comparison with one dataset (e.g. SOCATv5) additionally challenging.

Page 3 line 72-73: correct. Very little agreement was found in the SO, however, it is worth noting that Ritter et al also found remarkable agreement regarding decadal signals, despite the strong discrepancies in seasonality and IAV.

Page 9 lines 186-187: I find myself arguing a lot against the direct inclusion of spatial coordinates. The reason is that CO2 is not directly affected by longitude or latitude, but rather direct environmental proxies that vary along space and time (e.g. SST). Adding Lat and Lon might decrease the error metric as it replaces an unknown. However, in a process-sense it makes much more sense to apply some special selection by regimes, biomes or clusters.

Page 9 figure 3: I absolutely understand the advantage of adding additional regions to the “blank” Fay and McKinley biomes, but a bit more motivation would be good on how these additional regiosn were chosen (e.g. it is immediately obvious to combine the EBUS regions, but why not the very small Sea of Japan with the surrounding waters?)

Page 14 lines 322-323: This is a repeat and can be removed

Page 14 Figure 5 and following text: The authors in the text often use words like “outperform” (e.g. line 355, line 384) or “much larger” (e.g. line 364). As a reader when I hear outperform, I immediately think of a 50% error reduction or similar, whereas in fact not a single bias value in Figure 5 or table 3 qnd table 4 are above 1μatm. Just to put this into context: The current flag A measurement uncertainty is in the range of 1-2μatm. Hence these differences are small. They might be significant, but certainly don’t deserve wording like “outperformed”

Page 18 line 410: word “bias” occurs twice – remove first occurance

Page 24 lines 544-546: One has to be careful quoting the SOCOM study here. The SOCOM study faced 2 additional differences which made a combination trickier. Firstly, not all methods were based on the same observational dataset (some were built on 1 million data, others on almost 15 million at the time). Secondly, not all methods did span the same geographical extend (with some covering more of the coast, less of the Arctic, etc.). Based on all these differences (and others), Rödenbeck et al 2015 avoided to provide an ensemble mean estimate.

Page 24 and onward: The authors argue for the inclusion of EKE to move forward, but I was not fully convinced, given the small improvement in Figure 5. Another limiting factor that is not discussed is availability. In an ideal case one would use time varying fields of SST, SSS, DIC, TALK, etc. however, in most cases they are not available. As the authors mention, EKE only exists as a climatology, hence one cannot expect directly improved IAV signals from it (as e.g. visible in Figure 5c). I do however agree with the authors on their conclusions regarding the addition of novel proxies.

Page 26 line 611 –twice the use of the word “first”

Page 27 line 633: The authors name the Denvil-Sommer method as a way forward. I agree that this is an exciting new method, however – in the context of the text lines above – it is also not a fundamentally different method, as it is also based on machine learning.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-46, 2019.