Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-46-RC2, 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.

Jamie Shutler (Referee)

j.d.shutler@exeter.ac.uk

Received and published: 22 May 2019

The authors present a novel intelliegent interpolation method and a comprehensive analysis of current methods to intelligently interpolate pCO2 data (which then enable the calculation of global and regional atmosphere-ocean gas fluxes and net sink values). The analysis concludes that we are reaching the limit of improvement in these methods and that a increase in data coverage within those regions sparsely sampled is now needed to allow further advancement.

Whilst I agree with the final conclusion, I can see that the methods in the the analysis would benefit from further refinement to enable the 'wall' to be correctly identified. This refinement should focus on the consistent and more thorough handling of temperature within the data used for training and verifying the pCO2 interpolation methods. Improvements within the air-sea gas flux calculation itself would also enable a more accurate flux calculation.

Hence I would recommend revision of this paper before acceptance. This will require updating the methods and re-running the analysis.

Here are the major points that would need to be addressed. I suspect that addressing these temperature related issues will improve the overall results and conclusions.

The temperature issue relates to the section in the discussion that focuses on uncertainty. The strong temperature dependence of pCO2 means that a consistent handling of temperature is needed throughout the analysis to ensure that (inconsistent handling of) temperature is not the source of any biases and/or accuracy issues. This is especially true for a large collated dataset which has been generated from data collated from multiple systems, regions and ships (i.e. the SOCAT data). The SOCAT dataset is an amazing resource and vital in the study of air-sea CO2 exchange, but using it directly (in its original form) for this sort of global air-sea gas flux study can lead to unknown errors and biases. The need for correct temperature handling in air-sea gas exchange studies has been reviewed in depth by Woolf et al., (2016). The impacts of incorrect and inconsistent temperature handling can be significant, especially within large collated datasets and global analyses (please see the different examples and impacts that temperature can have on the different components of the gas flux calculation that are detailed within Woolf et al. 2016).

1. The SOCAT gridded dataset are based on gridded data that have all originated from different ships, systems and times. Hence the gridded values are likley to contain unknown biases due to inconsistent depths and thus temperatures that were originally

captured/linked to each pCO2/SST pair, but which was lost as a result of gridding (as SST and pCO2 are gridded individually). Furthermore, the differing depths of each sample means that multiple measurements within each box could be from different depths and so they are not part of the same (statistical) population. This issue can be overcome by first re-analysing the original SOCAT cruise data to a common temperature dataset (and thus a common sampling depth) and then re-gridding them (through the use of a satellite observed sea surface temparature dataset). This theory is explained in Woolf et al., 2016, the method is described in Goddij-murphy et al (2015) and the software tool to enable this is avialable within the FluxEngine open source toolbox (Shutler et al., 2016). A link to the github repository is below.

This needs to be done for the data used to train the methods and also data used to verify the performance of the methods.

2. The gas flux calculation itself as used by the authors (equation 2) is likely to add temperature related errors into the analysis. This bulk formulation using DpCO2 ignores vertical temperature gradients and so is likely to introduce additional (and unknown) errors into the analysis. Woolf et al., (2016) also discusses the shortfalls of using an inaccurate gas flux calculation. The work would benefit from using a version of the equation that accounts for differing solubilities at the top and bottom of the mass boundary layer. Section 2 of Woolf et al., (2016) provides the information needed to achieve this.

3. All of the independent methods used for the secondary validation also suffer from the issues of inconsistent temperature handling. This was a shortfall of the Rodenbeck et al. 2015. The work of Rodenbeck et al., 2015 was an excellent first step, but any verification using these data should consider the impact that inconsistent temperature handling is likley to have on any derived results. The authors should recognise this and discuss the issue.

4. The independent data used for verification (e.g. from GLODAPv2) will also suffer

from the same temperature issue. the GLODAPv2 data is also an amazing and very useful dataset. However, the data are all sampled from different depths and using different methods. So any conclusions drawn from deriving PCO2 from these data and its use for verification should consider the impact of inconsistent sampling depths and the resultant (unknown) temperature biases that this will introduce. The authors should recognise this and discuss the issue.

5. the GLODAPv2 data and its accompanying publication (Pfeil et al) provides an estimate of the bias within the GLODAPv2 data, which the authors here have used as an estimate of the uncertainty for the GLODAPv2 data. The bias will only be one component of the uncertainty (especially when using a large hisorical dataset like GLODAPv2 which spans many years, a period over which signifcant advances in methods have been made). The authors should mention and discuss this issue.

6. the GHRSSt dataset that is used within the analysis is (i think) a combined thermal and microwave dataset and so its valid for a specific depth (if mostly microwave data then its most likely sub-skin) making it valid for the bottom of the mass boundary layer (which is key when considering calculating the pCO2 and relevant solubility). if this is used as the input dataset for training the different interpolation methods then it would seem sensible to use this dataset for the gas flux calculation (to represent the temperature at the bottom of the mass boundary layer). This could also be used as the reference for re-analysing the SOCAT data prior to re-gridding. Doing all of this will ensure that your methods are consistent throughout.

Minor point, but still important: 6. Please can the authors clarify their descriptions of satellite observations? Satellite observations are not proxies (as stated on line 670). This is a common misconception, satellite observations are precise and accurate measurements of electromagnetic energy. e.g. some eatellite observations of SST are thermal infrared measurements and can be more accurate and precise than in situ sea surface temperature measurements (O'carroll et al 2008). The difference between satellite and in situ meaurements or obserations (apart from the method of collecting them) is predominantly the depth that the meaurement is valid for, as satellite sensors will retrieve the skin or sub-skin temperature, whereas an in situ measurement (which is a voltage measurement through a thermally sensitive resistor that is then calibrated to temperature) is normally collected at a few metres depth or somewhere near the surface. These differences are also briefly discssued in Woolf et al., 2016. For satellite chl-a its typically a visible spectrum measurement that uses a empirical relationship to calibrate the optical measurement to estimate chl-a. So the sentence (line 670) should be corrected to say 'satellite observations' (rather than proxy). please can the author check that the rest of their manuscript for further instances of the same issue?

References O'carroll et al (2008) Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations, https://journals.ametsoc.org/doi/full/10.1175/2007JTECHO542.1

Woolf DK, Land PE, Shutler JD, Goddijn-Murphy LM, Donlon CJ (2016). On the calculation of air-sea fluxes of CO 2. in the presence of temperature and salinity gradients. Journal of Geophysical Research: Oceans, 121(2), 1229-1248.

Goddijn-Murphy L, Woolf DK, Callaghan AH, Nightingale PD, Shutler JD (2016). A reconciliation of empirical and mechanistic models of the air-sea gas transfer velocity. Journal of Geophysical Research: Oceans, 121(1), 818-835.

Shutler JD, Land PE, Piolle J-F, Woolf DK, Goddijn-Murphy L, Paul F, Girard-Ardhuin F, Chapron B, Donlon CJ (2016). FluxEngine: a Flexible Processing System for Calculating Atmosphere–Ocean Carbon Dioxide Gas Fluxes and Climatologies. Journal of Atmospheric and Oceanic Technology, 33(4), 741-756.

FluxEngine software to perform the re-analysis and an accurate gas flux calculation can be found here: https://github.com/oceanflux-ghg/FluxEngine

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