Data-based estimates of interannual sea–air CO2 flux variations 1957–2020 and their relation to environmental drivers by Christian Rödenbeck, Tim DeVries, Judith Hauck, Corinne Le Quéré, and Ralph F. Keeling

The authors reconstruct historical air-sea CO2 fluxes from 1957-2020 using a mixed layer scheme constrained by observations of sea surface pCO2 from the SOCAT database.

The importance of SST, interannual variations in SST, and squared wind speed to the internal DIC flux from the mixed layer to depth are investigated using a multi-linear regression.

The multi-linear regression technique is used as a prior for a hybrid approach that estimates historical air-sea CO2 exchange from 1957 to 2020.

The authors’ methodology has clearly been exhaustively researched and many sensitivity studies have been explored to consider alternative options. This is work of a high quality that adds an important and new reconstruction of historical air-sea CO2 fluxes to the scientific field.

My comments/questions are:

The article would be improved if the sections outside of Method could have greater focus on the scientific findings instead of the method. For example, much of the discussion I expected to see in terms of comparison to other recent reconstructions was in the Appendix. This is quite interesting, and could be moved to discussion.

How do we know whether a “spin up” of 6 years (1951-1957) is adequate?

Why does including the final year (2021) cause problems?

Don’t we expect the sensitivity of the internal DIC fluxes to SST, wind speed, and interannual SST variations to be smooth because internal DIC fluxes are forced to be smooth within the approach?
The method is complex. The additional sensitivity studies are challenging to understand as presented in the main text. Moving these fully to supplementary would allow more emphasis on the main results. In the main text, it should be sufficient to state that the sensitivity tests have been done to explore X, Y, Z, and A, B, C are the primary things learned.

Some of the figures have many subplots, which takes away from the readers’ ability to see the main findings. A good number of the subplots – those not discussed in the text - could be moved to supplementary to make room for larger subplots of interest.

In the comparison in Figure 9 to other estimates, the text acknowledges the uncertainty on the river flux adjustment, but this uncertainty is not presented in the figure. Nor is the uncertainty in Fant from Gruber et al. 2019. If these were included, there would not be far less appearance of discrepancy between estimates. Instead of a dash line and a dotted line, a shaded area would be a better way to present this.

It would be helpful to add discussion of what other reconstruction approaches can learn from the findings here.

Page 22, line 21-26. Replace colons with periods here.

Page 22, last paragraph / Figure 10 and 11. Yes, an estimate of the pre-observed trend is needed, but it seems that this method is over-amplifying that trend. Can the authors propose approaches that might improve this going forward?