Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-339-RC1, 2022

Comment on bg-2021-339

Anonymous Referee #1

Referee comment on "A Numerical reassessment of the Gulf of Mexico carbon system in connection with the Mississippi River and global ocean" by Le Zhang and Z. George Xue, Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-339-RC1, 2022

Global climate models are of questionable utility in many regions due to poor spatial resolution and a poor reproduction of riverine inputs and other critical determinants of biogeochemical processes. Downscaling approaches are therefore critical in many regions. Zhang and Zhu present a new “downscaling” of CMIP6 model output for the region surrounding the Gulf of Mexico, and they draw conclusions about recent changes in the region’s carbon dynamics. The model used by Zhang and Zhu appears equally or more robust than prior models of the regional carbon budget. This is therefore potentially interesting and relevant work. However, in its present form the manuscript is needlessly confusing and misleading and features some potentially major methodological issues. I therefore recommend that the authors carry out a thorough revision of the manuscript text and to clarify methodological issues. The core contribution of this study is to provide updated (and potentially more robust) estimates of carbon fluxes in this region and to estimate temporal trends in variables such as pCO2 and pH. This is a valuable contribution to the literature as these values continue to have high uncertainties, and I hope the authors can address the concerns below.

1) It is highly misleading to call this a “downscaling” of a CMIP6 model.

At present, the title, abstract and introduction misrepresent the work in the paper.

The title of the manuscript claims this study downscales the global CESM2-WACCM-FV2 model. Conventionally, this should mean that all possible driving data is derived from the global model. Critically, any climate forcings should come from the global model. However, as stated on page 7 of the manuscript, the only things taken from the CESM2-WACCM-FV2 model are the initial conditions and boundary conditions on the geographic boundary. Atmospheric forcings etc. are not taken from the CESM2-WACCM-FV2 model. I therefore view this as a hindcast, where the authors were forced to use the

CESM2-WACCM-FV2 model for geographic boundary conditions as a compromise. In no real sense is it a downscaling of a CMIP6 model.

This is a major problem for the paper as there are, at present, many inaccurate statements. For example, the abstract claims this: “The model’s biogeochemical cycle is driven by the Coupled Model Intercomparison Project 6-Community Earth System Model 2 products (CMIP6-CESM2)…” This is clearly not true, as surface temperature, air PCO2, riverine inputs and most of the variables driving the carbon dynamics do not come from the CMIP6 product.

The title, and aims of the paper should therefore be revised.

The paper really appears to be a new estimate of carbon fluxes in the region. It should therefore be rewritten accordingly. Critically, the authors should make it clearer how, as claimed, the estimates in this study are more reliable than previous methods. The evidence provided for this are not extensive.

2) Use of the CESM2-WACCM-FV2 global model should be clarified

Output of the CESM2-WACCM-FV2 model are used for both initial and boundary conditions.

The authors do not state why they used the CESM2-WACCM-FV2 model for the boundary conditions. Was this model more accurate in the region than other CMIP6 models or reanalysis products that are available? This is a critical question, as it is possible the choice has reduced the reliability of the carbon budget estimates. There are also specific issues surrounding the use of this dataset.

First, this model can have negative values for nitrate, and presumably other variables. I viewed one of the historical files (http://esgf-data.ucar.edu/thredds/fileServer/esg_dataroot/CMIP6/CMIP/NCAR/CESM2-WACCM-FV2/historical/r1i1p1f1/Omon/no3os/gn/v20191120/no3os_Omon_CESM2-WACCM-FV2_historical_r1i1p1f1_gn_200001-201412.nc) for this model and negative values for nitrate appear very frequently across the boundary. Translating these values into boundary conditions is not a trivial issue as mass conservation etc. is ambiguous. The authors need to explain this thoroughly. Negatives at the boundary also result in average conditions that are far lower than those you would get from the NOAA World Ocean Atlas. Potentially this has been corrected for in some way by the authors, but if it has not it is not clear if the treatment of the boundary conditions is sensible. Likewise, there are negative values in the first time step in 2000, which the authors presumably used in some way to generate their initial conditions.

The authors state on p. 19 that this study's estimates of air-sea CO2 fluxes are "more reliable than previous GoM model studies". However, without showing whether the boundary conditions are reliable it is difficult to assess this claim. This is especially true, given the authors state that Xue et al. 2016 used over-simplified boundary conditions. There is therefore real potential that the boundary conditions used here are no more reliable than those in Xue et al.

Based on a comparison of this study with others, the approach to most variables is more robust than prior work, so the boundary conditions are likely the only major concern.

3) The model spin up period is potentially too short.

Only a single year is used for model spin up. It is not clear if the model will really have settled down by that point. Many regional models require 5 years to spin up, so one year is possibly questionable, especially given model output is used for temporal trend analysis.

Starting conditions are used from the CESM2-WACCM-FV2 model, and quasi-equilibrium conditions for this model will differ (perhaps quite dramatically) from the regional model. The authors justify using a one-year spin up by saying "the global model has been well stabilized up to the year 2000 from its 'pre-industry' experiment". This does not say much about the stability of the regional model used. Given the issues mentioned above about negative nitrate values in the global model, it seems questionable whether the starting conditions are close to a stable state in the regional model. Furthermore, it is plausible that riverine inputs are drastically better resolved in the regional model than the global model. This is particularly important given the conclusion of the importance of the carbon inputs from the Mississippi River.

The spin-up timing issue is also particularly relevant for the "no rivers" experiment. This experiment essentially removes rivers at the start of 2000, but assumes that the model is effectively spun-up to "river-free" conditions by the end of 2000. The authors need to show that this is credible. Otherwise, some of the results in the experiments section may not be robust.

4) Model validation needs to be improved

Overall, the model seems to do a reasonable job compared with observations. However, at present the model validation lacks rigorous statistics and is purely visual. There are 3 figures comparing model results and observations. However, there is a failure to show how close the model is to observations. I recommend the authors add correlation coefficients, RMSE and bias values for model-observation comparisons where relevant. These should give reasonable results based on the figures. This is particularly important for figure 5 comparing surface pCO2 between model and observation/ML model. The authors should also consider carrying out a similar analysis of pCO2 for the global climate model used to help assess the reliability of the boundary and initial conditions.

5) Figures should be made colour-blind friendly and made more clear

I recommend the authors ensure that all figures are colour-blind friendly. At least 7 of the figures are not. Figure 11 is very difficult to understand. Double y-axes should generally be avoided, and in this case they just serve to confuse. The axis units are also not stated.

6) Discussion and results should not be mixed up

At present, the results section includes discussion and the discussion includes results. Comparisons of the results with other studies (p. 19) should be moved to the discussion. Furthermore, the sensitivity analysis should be in the results section, not the results.

7) Temporal resolution of forcings should be clarified

The forcing data used is of varying temporal resolution, and some of it (such as oxygen) is only available as a climatology. The authors should clarify which driving data is actually changing during the 2001-2019 time period, and which are essentially unchanging. At present it is not fully clear what can and cannot be driving the temporal trends in carbon fluxes etc.

To what extent are the riverine inputs climatological? P. 7 states "Missing river alkalinity values are interpolated from climatological values, and missing river DIC values are calculated from pH and alkalinity…" An indication of how well varying riverine inputs are represented would clarify this.

The driving data sets mostly seems to be the best available, so minor clarifications are only needed.