Dear Editor and Reviewers:

Thank you for the 2nd round comments. We have tried our best to address these comments.

RC 1: *The authors have made major revisions to the manuscript based on previous comments by 2 reviewers. Overall, the authors have done a good job of addressing the comments of the reviewers. In particular, they have addressed the issue of the initial paper inaccurately describing the work as a dynamical downscaling of CMIP6 products. I believe the manuscript is almost ready to be published, but the authors should spend some time clarifying some elements.*

My comments will address a few aspects of the revised manuscript that the authors should address before publication. These comments focus mostly on new material in the manuscript.

Response: We want to thank both reviewers and the associate editor for the invaluable suggestions in the first round and second round review process, and we are clarifying these final elements in this comment.

RC 1: *Title*

The title is more accurate than the previous one. However, the use of the term "Reassessment" leads to an expectation that comparisons with previous studies would be more prominent, especially in the abstract. These comparisons are made in table 3, but they should also be included in the abstract. Overall, the study does show that the new analysis is likely more accurate than previous assessments, but it is probably not as clear as it could be. See further comments below.

Response: To better reflect the title, we added contents on regional model comparison in the abstract – "A reassessment of air-sea CO2 flux with previous modelling and observational studies give us confidence that our model provides a robust and updated CO2 flux estimation, and NGoM is a stronger carbon sink than previously reported."

RC 1: *Abstract*

"The biogeochemical boundaries were interpolated from NCAR's CESM2-WACCM-FV2 solution after a comprehensive evaluation of 17 Global Climate Model (GCMs) products against available observations and global climatology products". This statement is too strong for the abstract. The evaluation of the models is robust enough for this paper, but it is not particularly comprehensive in general.

Response: We adjust the statement to - "The biogeochemical boundaries were interpolated from NCAR's CESM2-WACCM-FV2 solution after evaluating 17 GCMs' performance in the GoM waters."

RC 1: *Description of model drivers*

The description of the model driving data has improved in the latest manuscript. However, it is still somewhat difficult to follow. I recommend that the authors state clearly in a couple of sentences what

their overall philosophy is for selecting the driving/boundary data, instead of only stating what data they have chosen. This will address obvious "Why did the study not use X?" questions from readers?
The authors should probably also state why they did not use a global biogeochemistry reanalysis product instead of a CMIP6 model. This appears justified, as I believe none of the openly available reanalysis have sufficient variables. But the choice of CMIP6 over a reanalysis may seem questionable to some readers, so clarification could help.

Response: In the previous submission (Line 201) we state, "The choice of CESM2-WACCM-FV2, among other GCMs, is primarily based on its horizontal resolution in the GoM region and its availability of nutrients and carbon variables (see Table A1 for more details)."

Table A1. Summary of CMIP6 GCMs considered for boundaries conditions of the regional model

Model Name	Institution*	Resolution (m) latitudinal × longitudinal	DIC	TA	NH4	NO3
CESM2	NCAR	54137×111951	available	available	available	not available
CESM2-FV2	NCAR	54137× 111951	available	available	available	available
CESM2-WACCM	NCAR	54137×111951	available	available	available	not available
CESM2-WACCM-FV2	NCAR	54137×111951	available	available	available	available
MPI-ESM1-2-LR	MPI	124664×124667	available	available	available	available
MPI-ESM1-2-HR	MPI	33395×42614	available	available	not available	available
MPI-ESM-1-2-HAM	HAMMOZ-Consortium	124664×124667	available	available	available	available
ACCESS-ESM1-5	CSIRO	109095× 99669	available	available	available	available
CMCC-ESM2	CMCC	97659×100093	available	available	available	available
CanESM5	CCCma	97659×100093	available	available	available	available
IPSL-CM6A-LR	IPSL	97659×100093	available	available	available	available
IPSL-CM6A-LR-INCA	IPSL	97659×100093	available	available	available	available
GFDL-CM4	GFDL	110769×99690	available	available	not available	not available
GFDL-ESM4	GFDL	110804×99690	available	available	available	available
NorESM2-MM	NCC	93221×99757	not available	not available	not available	not available
NorESM2-LM	NCC	93221×99757	not available	not available	not available	not available
NorCPM1	NCC	54137×111951	not available	not available	not available	not available

* Full name of Institutions:
CCCma: Canadian Centre for Climate Modelling and Analysis (Canada)
CSIRO: Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology (Australia)
CMCC: Centro Euro-Mediterraneo per I Cambiamenti Climatici(Italy)
IPSL: L'Institut Pierre-Simon Laplace(France)
MPI: Max Planck Institute for Meteorology (Germany)
NCC: Norwegian Climate Centre (Norway)
NCAR: National Center for Atmospheric Research (US)
GFDL: Geophysical Fluid Dynamics Laboratory (US)

In this revision we clarify the reason for choosing a specific as - "The two prognostic variables dissolve inorganic carbon (DIC) and total alkalinity (TA) are the essential data needed to drive a regional oceanic carbon model. There is no time-varying observational products or reanalysis of DIC and TA that has an ideal 3-dimensinal coverage of the GoM. NCAR's CESM2-WACCM-FV2 solution was chosen to serve as the model boundary due to its relatively small bias in the carbonate variables in the GoM, relative high horizontal resolution in the GoM compared with other GCMs and its availability of nutrients and carbon variables (see Table A1 for more details)."

The philosophy of using GCM to drive the regional model is 1) unlike some non-volatile elements in the ocean such as nitrate, iron, silicate, etc, carbonate system variables undergoes evolution and long-term trends with climate change under elevated atmospheric CO_2 forcing. The purpose of setting up a carbon model is at least partly to reflect the evolution of these climate sensitive gases and carbonate ion concentration in the water body. Using a reanalysis product without progression defeats the modeling purpose. Secondly, the connections between physical and biogeochemical variables are essential for understanding the ocean carbon cycle. Carbonate variables are not isolated but an indiscerptible portion of the mass and energy transport in the ocean. One key shortcoming of reanalysis products is the separation of these connections/processes. Extensive gap-filling techniques were applied to generate the climatology/reanalysis products when there is limited observational data. The preference between reanalysis products and GCM product is, in essence, a debate between non-mechanism and mechanism-based estimation model. For example, the raw TA observations from 2000-01-01to 2019-12-31 at all water depth including coastal estuaries in the GoM is 4189 counts (data included NCEI Accession 0083633, 0117971, 0144622, 0154383, 0157025, 0157461, 0157619, 0188877, 0188878, 0188879, 0188976, 0188977, 0188978, 0189038, 0189291, 0189592, 0208096, 0209158, 0219960, 0231438, 0240147, 0240177, 0240205, 0240206, 0240314, 0240320, 0240322), this already includes all publicly accessible TA observation data in the GoM. Assuming each observation data point is 100% representative of a 5km×5km grid point at its depth at the given month, when using the 36 vertical depth layout in this study, the observational data can only fill 2959 distinct grid points in the model. To generate a fully covered reanalysis product, substantial efforts need to be made in extrapolation if possible. Thirdly, introduced artifacts from data extrapolation is a problem from reanalysis products, such artifacts are random/method-bound and not supported by biogeochemical processes, and using a non-mechanism-based forcing for setting up a mechanism-based model can make the model assessment and optimization unreliable. For instance, bias from over-upwelling or over-calcification may be mixed with extrapolation artifacts, this makes model diagnose and process optimization elusive. It should be noted that empirical relationships with temperature and salinity are widely employed for DIC and TA extrapolation and in climatology product preparation (e.g., Xue, et al. 2016).

The only two climatology products containing both DIC and TA for the global oceanic waters are "OceanSODA-ETHZ" (Gregor and Gruber, 2020) and "A global monthly climatology of total alkalinity (AT) and total dissolved inorganic carbon (DIC)" (Broullón et al., 2020a,b). However, none of them can provide three-dimensional DIC/TA data with time progression, with the former covering from 1985 through 2018 by taking advantage of satellite data yet only contain surface information and the latter being a 12-snapshot monthly product erased time component. We explain the reason for not using a reanalysis product concisely in the revised manuscript as well.

RC 1: *Figure 15*

I recommend that the authors either remove figure 15 or reconsider the colour scheme. It is currently very difficult to make out where the model is positively or negatively biased. I can see that there is a large *bias near the coast, but elsewhere I cannot easily tell if there is a positive or negative bias. This could be fixed by using a diverging colour palette.*

Response: Thanks for the suggestion, we originally used a non-diverging color palette to distinguish the blank background and data points what has small bias. We have changed into a diverging color palette with the neutral color being the same as the blank background, and only included the regional model results. We added a statement that the blank background color does not indicate neutral bias in the figure caption to prevent confusion.

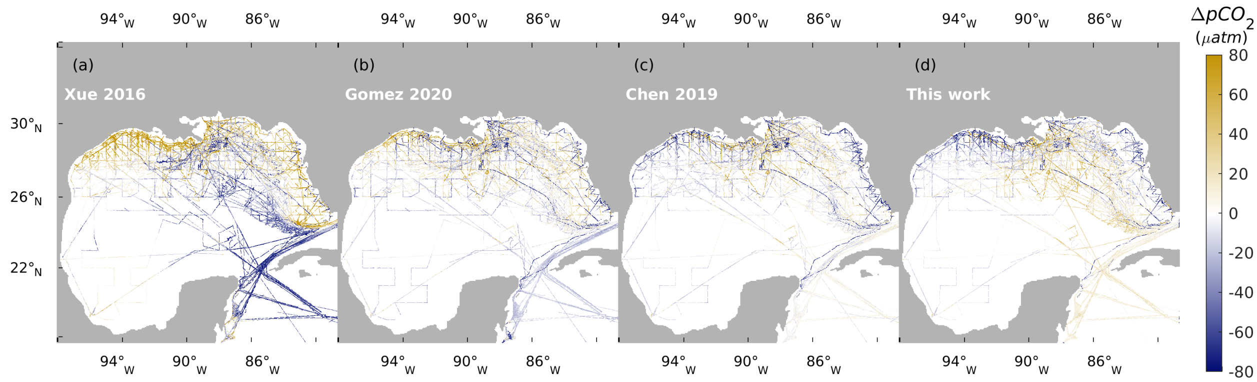

Figure R1-1 (Figure 15 in new revision): Comparison of sea surface pCO_2 between regional ocean model products (Xue 2016, Gomez 2020, Chen 2019, This work), and underway sea surface pCO_2 measurements. A Positive ΔpCO_2 indicates the product data overestimate sea surface pCO_2. A negative ΔpCO_2 suggests the product data underestimate sea surface pCO_2. A neutral ΔpCO_2 indicates the product data agree well with the observed sea surface pCO_2. The white spaces between the cruise lines indicate these regions do not have observational pCO_2 data, and do not indicate neutral bias.

RC 1: *Figure 16.*

This is an importamt and useful figure, in that shows the better performance of this model versus the others assessed. However, the authors should consider making some changes.

I would not expect the coarsely gridded CMIP6 models to be particularly good at resolving pCO2 in this region. It should be easy to outperform them, and so the comparison with CMIP6 should probably go into the supplementary materials.

The important comparison seems to be with the previous regional models. I therefore recommend redoing figures 15 and 16 to only include the regional studies. This is also more in line with the "Reassessment" aspect of the title. It is important the authors make totally clear how their study is improving on previous ones, and figures 15 and 16 weaken that aspect of the study.

Response: Thanks for the suggestions for Figure 15 and 16, we modified the figures to only include the regional studies to be more in line with the "Reassessment". And the updated figure is also shown below.

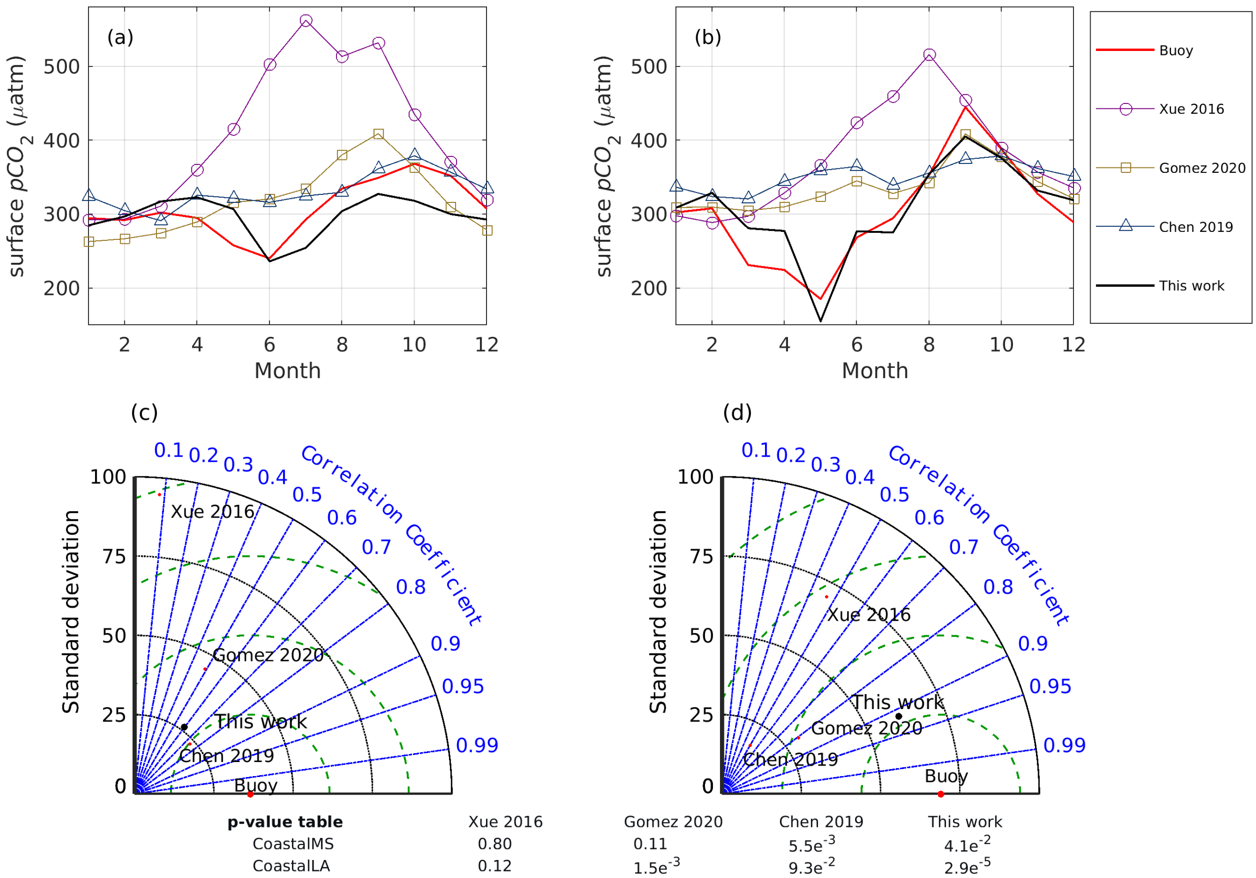

Figure R1-2 (Figure 16 in new revision): Comparison of sea surface pCO_2 among regional ocean model products (Xue 2016, Gomez 2020, Chen 2019, This work) at two buoy sites. Climatology at the two buoy locations of Gomez et al. (2020) is calculated by multiyear averaging from 2000-2014 model surface results. Climatology at the two buoy locations of Xue et al. (2016) is calculated by multiyear averaging from 2005 to 2010. Climatology at the two buoy locations of Chen et al. (2019) is calculated from their 12-monthly ML surface pCO_2 product (from 2002-07 to 2017-12). Buoy raw observations have a frequency of ~ 3 hours, and monthly averages are used to be compared with monthly model estimates. The p-value for each correlation coefficient is listed in the p-value table.

RC 1: *Table 2*

As above, the authors should consider restricting this to the regional models. Ideally, table 2 will only include the models shown in table 3. There is also a potential technical issue with the calculations for the CMIP6 models. Many of the coastal points will be outside the CMIP6 model domains. It is therefore unclear how they were handled. Were they extrapolated outside the domain? Given the heavy concentration of coastal points, it is possible the comparison between CMIP6 models and this study is not consistent..

Response: We reorganized Table 2 to only include the models shown in Table 3. Thanks for pointing out the potential technical issue of directly comparing the observation with CMIP6 models. We have included a similar statement in the original manuscript that "(Line 48) However, their relatively coarse spatial resolution is likely not appropriate to be directly compared with field measurements". This is also a reason to develop a regional model to better fill the gap between observations and models. In our first submission, we did not perform a such direct comparison between GCMs and observations, and it was added to provide more information upon request. Due to the coarsely resolved coastline, some coastal observation data points might be outside of the CMIP6 model domain. In such cases, we have to force the CMIP6 model to give its best estimation from the nearest data points of the nearest timespan. Indeed, observation data represent a much smaller geographical area over a limited time span. It is unfair for the CMIP6 models to be compared with the regional model or observations, in that they have different horizontal resolutions and time-frequency. However, the results are the best estimates CMIP6 monthly data products can provide. An extrapolation from the nearest data point is better than no estimation, if there were no other choice, global model products can still be a good source for making a such estimation. The product comparison is not consistent as limited by each product's capability, but it is consistent in that all products are allowed to give their best estimations.

RC2: *Review of "A Numerical Reassessment of the Gulf of Mexico Carbon System in Connection with the Mississippi River and Global Ocean"*

This paper presents a high-resolution model of the Gulf of Mexico with a fully coupled marine carbon cycle included. It demonstrates that the model performs well and better than climate models in this region with respect to the variables evaluated. Furthermore, the model results are used to investigate trends and changes in the surface carbon variables and exchange with the atmosphere over the simulation period (20 years). The importance of lateral fluxes and riverine input to the region is also investigated. This paper makes a substantial contribution to the science of biogeochemical cycling in the Gulf of Mexico. This is the second time I review this paper and I think that my initial concerns have been comprehensively addressed and I therefore recommend the paper for publications with a few minor corrections.

The biggest request for change is that I think large parts of section 5.1, equations, figures, results fit better at the end of section 3 than in the discussion section. A small part where you discuss the differences between the models can be retained under 5.1.

I appreciate the fact that so few observations go into the climatology close to the boundary as shown in figure R1-1, but I think the units must be wrong in this figure: should it be micromoles/kg?

A couple of typos:
Line 16: should it be "generally interesting carbon …"?
Result, line 452: Indice should be index

Response: Thanks for the careful review of this manuscript and the invaluable suggestions in both rounds of comments. We have re-examined the layout of section 5.1 and find the discussion content involving multiple global climatology products and global models necessary. Thus we think it is reasonable to leave it as current section rather than being moved to the end of validation section (section 3).
The unit for Figure R1-1 should be *micromoles/kg*, sorry for this oversight. A corrected figure is attached below as Figure R2-1.