Review of "***Linking global terrestrial CO_2 fluxes and environmental drivers using OCO-2 and a geostatistical inverse model***" by Zichong Chen *et al*.

This study presents the linkages between flux estimates derived from OCO-2 retrievals and environmental drivers across globally 7 biome-based regions. Using a geostatistical inverse modeling (GIM) approach, the authors demonstrate that they are able to identify connections between carbon flux and three key environmental drivers, namely air temperature, daily precipitation and PAR – the combination of these three variables explaining more than 89.6% of the variability in CO_2 fluxes. However, the study is conducted for one year only (circa 2016). The authors claim that this is an initial case study, thus implying that a more comprehensive study will follow later. This begs the question – is this study intended to demonstrate that the GIM approach has been successfully adapted to remote-sensing observations (i.e., a technical study) or is it intended to capture the connections between CO_2 fluxes and environmental drivers (i.e., a scientific study)? Kindly see Major Comment #1.

I believe the authors ideally wanted it to address a bit of both but unfortunately, in trying to address both, the authors end up addressing neither. I highly recommend that the authors take a step back and decide whether to focus on the inversion methodology and application to OCO-2 retrievals **OR** highlight the scientific questions related to regional and seasonal environmental drivers, and then resubmit. In general, the manuscript is well-written and concise, but it falls short of a clear formulation in terms of scientific scope, depth and novelty.

Several other questions persist. These revolve around limited validation of the posterior flux estimates or posterior CO_2 concentrations (see Major Comment #4). The choice of the model-data-mismatch variance (**R**) is inconsistent with real OCO-2 retrievals and needs justification in the main text (rather than bypassing it and relegating it to the Supplementary Section). **R**, along with the a priori flux covariance matrix **Q**, balances the relative weight of the atmospheric data and the trend in estimating the fluxes. An inverse modeling study cannot gloss over these details (see Major Comment #6).

Along with my comments below, I have suggested a few basic analyses and additional experiments, that will improve this study and make it scientifically robust and appealing to the larger carbon cycle science community. I sincerely hope that the authors consider these suggestions.

Major Comments:

1) *Scope of the study* – as mentioned earlier, the authors need to lay out a clear scope early on and remain consistent throughout. If the authors are interested in examining the relationship between carbon flux and environmental drivers, a one-year study is not justifiable. The authors need to examine the relationship over a number of years, make sure they are capturing the inter-annual variability in their flux estimates and then assess the relationship between drivers and fluxes. In addition, it is worth noting that the selected year is an El Niño year. On Page 3, Lines 86 – 88, the authors justify this decision by pointing out that the OCO-2 observations had 7-week gap in 2015- and 1.5-month gap in 2017. Remote sensing datasets, or rather any real observations, will always have data gaps! Simply discarding entire years' worth of data for a 5-7-week gap is not a reasonable justification. On the other hand, if the authors want to highlight the development of a new inversion framework/methodology, then it may be out of scope for ACP, and may be better suited to a journal like GMD, where a

lot of the mathematical nuances can be captured. Right now, a lot of the important mathematical details have been relegated to the supplemental material, including important discussions about the error covariance parameters and how they are derived. These details need to be included in the main text.

2) *Scientific novelty* – The authors report that a combination of PAR, daily temperature and daily precipitation are the most adept at capturing the variability in the fluxes (PAR for mid-to-high latitudes and a combination of daily temperature and precipitation for the tropical biomes). Neither of these findings are unique. The authors have correctly referred to a host of studies using GIM (e.g., Gourdji et al. 2008, Fang and Michalak, 2015, among others) or studies using OCO-2 data that have examined the response of the land carbon cycle during the 2015-2016 El Niño (e.g., Liu et al., 2017, Crowell et al., 2019). The BIC did its job and picked up the variables it was supposed to; hence, it is slightly unclear how this study adds new insights into our knowledge about carbon cycle science. In fact, by the authors own admission in Sections 3.1.1 and 3.1.2, almost all their findings are exactly the same as reported in previous studies. These two sections almost read like a literature review rather than a results section with new and exciting science results.

3) *Selection of auxiliary variables and how they are being reported* – what may add a new dimension, relative to already published studies, is reporting a table with all the 12 selected environmental drivers and including the estimated drift coefficients, coefficient of variation, annual average contribution to flux and the correlation coefficient between the selected auxiliary variables in the model of the trend. Actually, the annually averaged global contribution to flux can be reported in typical carbon flux units (like GtC/yr or PgC/yr). That would be novel information, especially if it were to be compared against estimates based on in situ data. Finally, just out of curiosity, why didn't the authors select *fPAR* instead of *PAR*? Also, the authors argument for not including LAI or SIF because they are "remote sensing indices" (Page 5, Lines 144-146) is surprising. Almost all of the auxiliary variables listed on Lines 138-141 are derived from remote-sensing measurements. What if the authors were to include LAI? How would that change their selected model of the trend?

4) *More rigorous evaluation of posterior flux estimates and more importantly, posterior concentrations, against independent measurements* – The biggest surprise of this study is that there are extremely limited evaluations presented against independent measurements (only 7 aircraft sites!). Given the large number of available independent datasets (*in situ* such as surface flask sites, towers and aircraft, TCCON), the absence of a detailed evaluation is striking. Especially, from a seasoned inverse modeling team. Since the authors claim that they are estimating daily global CO_2 fluxes at the GEOS-Chem grid scale (Page 3, Lines 72-73), there should be no reason for not evaluating against observations from dedicated aircraft campaigns such as ATom or ACT-America. In addition, it is also not clear why in Section S7, the authors allude to the results from Crowell et al. 2019. The authors have to back up their own biases and RMSD and explain those numbers and their significance, rather than pointing the reader to Crowell et al. 2019 for justification.

5) *Comparison of findings against those derived from in situ data* – The value of this study will be significantly enhanced, if the authors do the same analyses utilizing in situ data (such as

NOAA obspack). Are the conclusions, especially in terms of the three significant drivers and their contribution to the carbon flux, consistent? It has been 12+ years since the Gourdji et al. 2008 study attempted such an analysis – given the increase in the number of surface flask sites and improvements in atmospheric transport model, availability of auxiliary datasets, it will be worth revisiting this and comparing against the information reported here from OCO-2 datasets.

6) *Error covariance parameters* – Can the authors explain why they switched to a spherical covariance model instead of sticking with a simpler exponential covariance model? The authors argue that the shorter correlation length is due to higher density of observations relative to previous studies. Part of that is true. But I believe that the shorter correlation length in the residuals is more reflective of the model of the trend that has been fitted to large biome scales. The model of the trend is too complex for the biome scale; for the grid scale studies that the authors allude to, it made sense. Additionally, the authors persist with a model-data mismatch variance of 1.19 ppm^2 based on a previous pseudo-data study. Why? I highly encourage the authors to use the reported X_{CO2} uncertainty for the OCO-2 soundings and then add reasonable representation of transport and representation errors to get 'real' MDM variances. This shouldn't be a huge task given the involvement of core GEOS-Chem developers in this study. It wouldn't be surprising if more reasonable **R** values lead to an increase in *a posteriori* uncertainties for their flux estimates (Page 11, Lines 324-325).

--
Abhishek Chatterjee
Global Modeling and Assimilation Office (GMAO)
Goddard Earth Sciences Technology and Research (GESTAR)
NASA Goddard Space Flight Center
Mail 610.1 | Greenbelt, MD 20771 | USA

Phone: +1 (301) 286-7870
Fax: +1 (301) 614-6246
Email: abhishek.chatterjee@nasa.gov
Url: https://sciences.gsfc.nasa.gov/sed/bio/abhishek.chatterjee