We thank the reviewers for their detailed suggestions and comments on the manuscript. We have re-written the manuscript, added substantial new analysis, and included extensive new comparisons against independent observations based upon the reviewer suggestions. Below, we have replied to each review and have detailed the corresponding edits that we have made to the manuscript. We have listed out the reviewer comments in italic font and the replies in regular font.

RC1: Referee #1

The authors have developed a geostatistical inverse method to interpret satellite observations of carbon dioxide (CO_2) collected by the NASA Orbiting Carbon Observatory collected during 2016. As far as this reviewer can see the study is scientifically sound but describes only an incremental improvement to the method and does not lead to any new scientific insight.

We have re-written most of the manuscript, overhauled the inverse modeling setup, and added substantial new analysis to improve the novelty and scientific messaging. Specifically, in the revised manuscript, we have added the following new analyses:

- We compare the environmental relationships that we infer from OCO-2 against the relationships that we infer from 15 terrestrial biosphere models (TBMs) from the recent TRENDY model comparison project (Sect. 3.3).
- We evaluate when and where TBMs agree and disagree on these relationships and what factors might be driving these disagreements among TBMs (Sect. 3.3).
- We have expanded the analysis from one year to four years.
- We have added analysis to better explore what factors limit our ability to infer these environmental relationships using current satellite observations from OCO-2 (Sect. 3.1).
- We have added extensive evaluation against ground-based CO_2 observations (the Supplemental Sect. S4, Figs. S2-S12, and Tables S2-S3).

The environmental drivers for ecosystems located at mid/high and tropical ecosystems are unsurprising. Perhaps that's the point. I wasn't sure. PAR is by definition photosynthetic active radiation so its ability to describe large-scale CO_2 fluxes isn't anything new, particularly over one year that is dominated by the seasonal cycle. Any insights from using the diffuse and direct components of PAR? Similarly, temperature and precipitation roles in the tropics are nothing new. However, I am surprised that precipitation is such a useful driver over the tropics where complex basin-scale hydrologic controls are at play. In other words, where it rains is not necessary where the water ends up.

We agree that these environmental drivers are unsurprising. In the revised manuscript, we have added analysis comparing the relationships that we infer from OCO-2 observations against those inferred from 15 TBMs (Sect. 3.3). Existing terrestrial biosphere models (TBMs) disagree on the relationships between these environmental drivers and CO_2 fluxes; TBMs show a large range of relationships, and for some variables like precipitation, TBMs often disagree on the sign of that relationship. We feel that this new comparison with process-based models provides better depth and novelty to the manuscript.

It is true that where it rains is not necessarily where water ends up, particularly at fine spatial scales like the scale of a stream catchment. In this study, we model fluxes at a much broader

spatial resolution that reflects the resolution of the GEOS-Chem model (4 degrees latitude by 5 degrees longitude). At that broad scale, patterns in spatially-averaged precipitation are more strongly correlated with surface soil moisture than at finer spatial scales. Note that we ran several test simulations where we offered up both precipitation and soil moisture as auxiliary variables in the inverse model, but the model selection framework only chose one of the two (precipitation); those two predictor variables were highly colinear or correlated, indicating that the inverse model did not have the power to distinguish between the two. Furthermore, precipitation was included as a standardized input variable in the TRENDY model inter-comparison, so we wanted to at least offer up precipitation as a candidate auxiliary variable in the analysis of OCO-2 and the TRENDY models.

The authors have gone some way to 'fess up that the geostatistical inverse method uses prior information for which I commend them. It might not be defined in the same way as the classical Bayesian approach but nonetheless it uses prior information. Otherwise, inferring fluxes for 10^6 grid boxes using 10^5 measurements is an ill-posed problem. The method uses environment driver data with uncertainties that are difficult to quantify (see comment below about estimated posterior uncertainties).

A geostatistical inverse model certainly does use prior information. That information is just in a different form than other types of Bayesian inverse modeling.

It would be useful to reiterate to the reader the benefit of the geostatistical inverse method over more traditional methods. Certainly, it provides an alternative perspective but I have seen no evidence to suggest it is better or worse.

In the revised manuscript, we have added substantial new analysis to better highlight new insights facilitated by this approach. This new analysis includes a comparison of the environmental relationships that we infer from OCO-2 against those inferred from 15 state-of-the-art TBMs (Sect. 3.3). Existing studies have used this geostatistical approach to compare the environmental relationships in different TBMs (e.g., *Huntzinger et al*. 2011) and to compare with the relationships inferred from in situ atmospheric observations (e.g., *Fang and Michalak*, 2015). In the revised study, we build upon that existing body of work by comparing the relationships inferred from OCO-2 across the globe with those inferred from TBMs.

Line 216: This reader is surprised that OCO-2 data are not sensitive to biomass burning emissions, particularly during the El Nino period. The manuscript would benefit from having more explanation on this point.

We have overhauled the inverse modeling setup to include more prior information on biomass burning (from GFED) and ocean fluxes (Sect. 2.6). We have also added a new discussion in the results (Sect. 2.6) and SI (the Supplemental Sect. S2) describing the contribution of biomass burning fluxes relative to other types of fluxes. In these sections, we also discuss why biomass burning fluxes are challenging to uniquely identify and constrain in an inverse model. Specifically, the atmospheric signal from biomass burning (as estimated by GFED) is small (0.19 ppm) relative to anthropogenic emissions (2.7 ppm) and model-data errors specified in the inverse model (standard deviation of 0.29 ppm to 4.8 ppm).

Why are correlations higher when environmental drivers are passed through the atmospheric model. Figure 3 doesn't cut it - the color scale is almost binary as currently defined. Using the square of the correlation might be a better way to illustrate these calculations.

We agree that Fig. 3 in the original manuscript was confusing and have re-designed this discussion to more clearly communicate the message we intended to communicate. We have replaced Figure 3 in the revised manuscript.

Line 263: widespread and prolonged drought conditions, together with large-scale land-use change, is a more accurate description of what's going on over these regions.

Noted. We have edited Sect. 3.3 in the revised manuscript accordingly.

Paragraph 298: comparison of the reported work and other groups is weak. Not many people have used v9 of OCO-2 data so I think it would be useful for the readership to provide a more detailed assessment of results compared with past estimates using v7 data. The comparison between the model and independent measurements is minimal (in supplementary information). The uncertainties associated with the posterior estimates are unrealistically small. The classical Bayesian inversion as typically employed underestimates posterior uncertainties so certainly the uncertainties estimates reported with the geostatistical method are grossly underestimated. This reviewer is left wondering why this might be so and how a possible explanatory imbalance between prior and observation uncertainties would influence model selection and the analysis that follows.

We have added substantial comparisons against independent measurements in the revised manuscript, including comparisons with observations from twenty regular aircraft sites (the Supplemental Sect. S4; Figs. S3-S6), the Atmospheric Tomography Mission (ATom) (Fig. S7), and 18 sites from the Total Carbon Column Observing Network (TCCON) (Figs. S8-S12). Note that, in the revised manuscript, we have used version 9 for the analysis because version 7 observations are now several years outdated and contain much larger observational errors.

We have also compared our results against the most recent provisional results from the inverse modeling inter-comparison (MIP) project that uses version 9 of OCO-2 retrievals (refer to the figure below). We find that our flux estimate is usually close to the ensemble mean of the v9 MIP and is always within one standard deviation of the MIP estimates. Note that the results shown below from the v9 MIP are from the MIP website (*https://www.esrl.noaa.gov/gmd/ccgg/OCO2_v9mip/*) and are provisional results that have not yet been finalized.

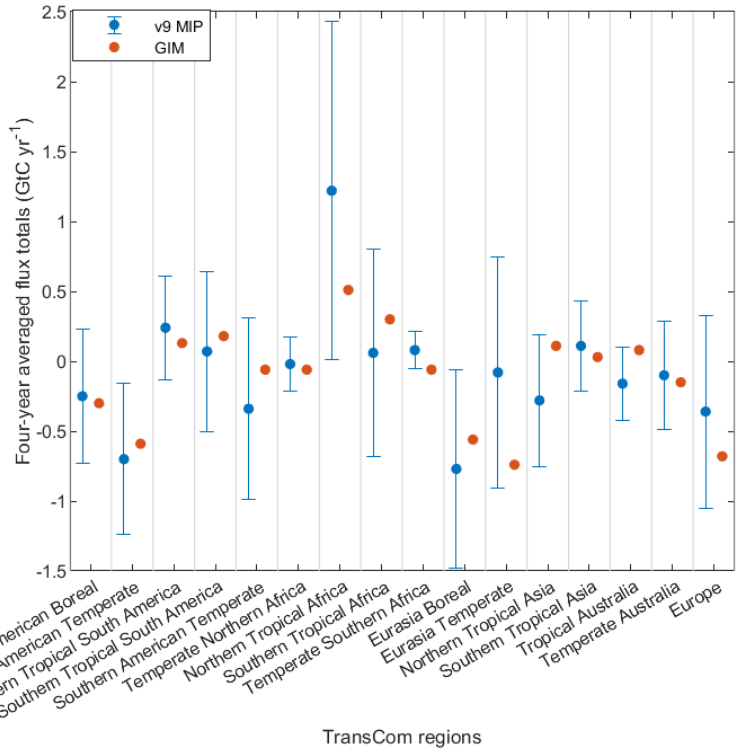

Figure R1. Comparison of biospheric flux estimates by TransCom region from this study (red) and the v9 MIP (blue). Error bars in the MIP results indicate one standard deviation of flux estimates across the ensemble. Our best estimate is usually close to the mean of the v9 MIP study and is always within one standard deviation of the MIP results. Furthermore, the GIM estimate does not show any consistent bias relative to the MIP ensemble mean.

We have also overhauled the inverse modeling setup and have set improved values for the covariance matrices in the inverse model (**R** and **Q**) (Sect. 2.4 and the Supplemental Sect. S1). For example, the model-data mismatch errors are now based upon the reported errors in the 10-second average OCO-2 data product. We have further estimated the relationships between CO_2 fluxes and environmental driver datasets using two different meteorological products (MERRA-2 and CRUJRA) to explore the sensitivity of these results to the choice of meteorology used for the driver datasets. We believe that this revision has yielded better uncertainty estimates in the revised manuscript.

Sure, the tropical flux estimates are important to discuss. However, are the reviewers are in a position to dismiss the results over tropical North Africa without further explanation. Why did they find themselves in terms of environmental drivers? Surely, their results over tropical Africa aren't exclusively determined by measurements collected over tropical Africa? Do they find that seasonal differences in measurement over tropical Africa lead to a bias in the flux? Answers to these questions would represent a useful contribution to the field.

There is almost nothing in the manuscript about the large differences between other geographical areas where we would expect much better agreement, e.g temperature North America, Europe, Eurasian temperate. Without a more comprehensive evaluation of the fluxes it is difficult to know whether the method is at fault or the data they have used. This manuscript would benefit greatly from a better evaluation of the posterior fluxes.

We agree that comparing our results using version 9 of the observations against studies that used version 7 is not necessarily a fair comparison; there are large differences between v7 and v9 of the OCO-2 observations, and differences between existing studies using version 7 and our results using version 9 could reflect differences in the observations as much as differences in inverse modeling methodology. When we compare the GIM flux estimate against provisional results from the most recent MIP, we find much better agreement between our results and the MIP; our estimate is always within one standard deviation of the MIP ensemble mean.

In the revised manuscript, we also evaluate our inverse modeling results using numerous ground-based datasets (the Supplemental Sect. S4, Figs. S2-S12, and Tables S2-S3). We feel that these new model-data comparisons provide a much-improved evaluation of the posterior fluxes.

Line 59: it would be fairer that Chevallier 2018 argues not suggests.

We have edited this line accordingly.

For context, it would be useful for the reader to understand that 2016 was an El Nino Year.

We include four years of observations in the revised manuscript (instead of the one year in the original manuscript). We also point out in Sect. 3.3 that 2015-2016 are El Nino years.

Line 91: how did the authors decide that four months was a sufficient spin-up period?

We have clarified this point in the revised manuscript (the Supplemental Sect. S1). We used this setup for the model spin-up because it is the same setup used in *Miller et al*. (2018). We first created an initial condition for 1 Sept., 2012 based on NOAA's Carbon Tracker (CT) product, and used CO_2 fluxes from CT to run GEOS-Chem forward for two years until 1 Sept., 2014 when the inverse modeling begins; we ran the CT fluxes through GEOS-Chem for two years to make sure the CO_2 mixing ratios are consistent with the GEOS-Chem model grid, and therefore to minimize potential spin-up artifacts due to model transport. We then run the inverse model starting from 1 Sept., 2014, but we consider the result from 2014 as part of an initial model spin-up period and do not use it for analysis.

SC1: Brad Weir

In this work, the authors use a geostatistical inverse modelling approach to infer surface fluxes from observations of column CO_2 by the Orbiting Carbon Observatory 2 (OCO-2). Using these estimates, the authors make claims about the environmental drivers of the spatiotemporal variability of surface fluxes. However, their evaluation against independent data (sometimes coarsely defined as "validation") is not sufficient to support these claims.

We have added extensive evaluation against ground-based observations, including from 20 regular aircraft sites (the Supplemental Sect. S4; Figs. S3-S6), the Atmospheric Tomography Mission (ATom) (Fig. S7), and 18 sites from the Total Carbon Column Observing Network (TCCON) (Figs. S8-S12). We find that model-data biases are small across most of the globe (except at sites near urban regions) and that the standard deviation of the model-data residuals is within the uncertainties specified in the inverse model (i.e., is within the model-data mismatch specified within the covariance matrix). The Supplemental Sect. S4 of the revised manuscript includes a detailed discussion of these model-data comparisons.

We have also compared our flux estimate against provisional results from the most recent OCO-2model inter-comparison (MIP) project, and our flux estimate is typically close to the ensemble mean and always within one standard deviation of the mean (refer to Fig. R1 above).

Inferring surface carbon fluxes from observations of atmospheric CO_2 is an inherently ill-defined problem. Its solution, in any form, requires a number of assumptions that are often poorly constrained by existing scientific knowledge. The authors do a commendable job of explaining that despite erroneous claims in the existing literature to the contrary, geostatistical inverse models do in fact use prior information, just in a different form than more common approaches. What the authors fail to do is support that their surface flux estimates are fit for the scientific purpose at hand. Typically, this is accomplished through comparisons to other independent data products. While pedantic, it seems more and more necessary that we remind ourselves that inferred surface fluxes fall into the prediction step of the Scientific Method. Between that and the analysis step, is the all important testing step. The testing step cannot be shortcut – it is the only thing separating science from plausible guesswork.

In order to make claims about the spatiotemporal variability of surface fluxes, the authors must first evaluate the fidelity of their surface fluxes' spatiotemporal variability. While this reviewer admits that there is no ideal method of evaluating global surface fluxes of CO_2 on horizontal scales greater than a few tens of kilometers, a greater effort must be made to demonstrate the product is appropriate for the analysis in the text. In particular, the only evaluation of their surface fluxes is that of long-term time mean regional budgets (Figure 6) and simulated CO_2 at just a handful of aircraft profiling sites (Figures S5 and S6). If one is to make claims about seasonal cycles, for example, then the seasonal cycle of the inferred fluxes must be evaluated as well. Given the assumptions necessary to make these inferences, it is entirely possible that their long-term time mean budgets are reasonable and their seasonal cycles are not. This is especially important given the documented impact (Basu et al., 2013, ACP; Crowell et al., 2019) that very small seasonal and regional biases from satellite retrievals can have on inferred fluxes. Unless the authors are able to demonstrate the skill of their product in reproducing variations over the

same spatiotemporal scales as the scientific analysis, this review does not see how their claims can be supported.

Thank you for the suggestions. We have greatly expanded the model-data evaluation in the manuscript (the Supplemental Sect. S4; Figs. S2-S12; and Tables S2-S3). In the original manuscript, we compared against a handful of aircraft sites, as the reviewer points out. In the revised manuscript, we compare against numerous additional aircraft sites, as well as comparisons against TCCON, and comparisons against campaign data from ATom.

RC2: Julia Marshall

At first glance it seems that the results of this study make sense, and are consistent with our general understanding of what drives carbon fluxes, with uptake at higher latitudes being mostly radiation-limited while in the tropics there are more complex temperature-precipitation interactions. So far, so good. The paper is well written and clearly structured, making it easy to read. To the careful reader it soon becomes clear that something is going wrong, however, and the limited "validation" and comparison to other results from the literature are insufficient to explain these problems away. While the geostatistical approach is com mendable in that it allows more flexibility in the structure of the prior fluxes, such that perhaps unexpected signals may emerge, it also seems to allow for rather unphysical results, as in this case. Given the fact that the ocean fluxes (a net sink of more than 2 PgC/year) were rejected by the Bayesian Information Criterion (BIC) while the net land fluxes are more or less consistent with other studies, it seems impossible that the global atmospheric growth rate can be matched. It just does not add up.

We estimated ocean fluxes alongside terrestrial fluxes in the inverse model in the original manuscript but did a poor job of communicating those results. In the revised manuscript, we have not only improved the discussion of ocean fluxes but have also overhauled the inverse modeling setup to include more detailed prior information for ocean fluxes. In the revised manuscript, we use prior information for ocean fluxes from the NASA Estimating the Circulation and Climate of the Ocean (ECCO) Darwin flux product. In our original setup, prior ocean fluxes from Takahashi were not selected using the BIC, and the inverse model instead defaulted to a non-informative prior over the ocean. In the revised setup, we have grouped together ECCO-Darwin, anthropogenic emissions (from ODIAC), and biomass burning emissions (from GFED) into a single column in the auxiliary variable matrix (**X**). ECCO-Darwin, when included as a separate column of **X** is not selected, but a column of **X** that includes all of these prior emissions estimates together is selected.

We describe this updated setup in Sect. 2.6 and the Supplemental Sect. S2 of the revised manuscript and show ocean fluxes alongside terrestrial fluxes in the inverse modeling results in Fig. 5.

This should be obvious when performing validation, but the very little testing of the posterior fluxes, limited to a handful of aircraft measurements far from coasts on a scatter plot averaged (monthly?) by height, hidden in the supplement, makes it hard

to tell. The paper states that aircraft profiles near coasts were not used because the coarse model resolution made it hard to represent these data well, but I wonder if the complete absence of ocean fluxes may have also played a role here?
Since none of the in-situ sites were used for constraining the fluxes (which seems an odd choice, even if only for comparison's sake), it would be instructive to plot the concentrations resulting from the posterior fluxes at a few sites to see if the curves drift apart over the year as a result of the missing sink. While this might not look too bad in a simulation of only one year, this would soon result in wildly divergent curves. But perhaps over a longer simulation the BIC would then choose to select the ocean fluxes. Still, the decision to blindly allow the model to return what we know is incorrect makes it hard to trust the interpretation of the results. Perhaps Takahashi was not the best ocean prior in this case, especially for an El Niño year, and this played a role: this could be an area for more analysis.

In the revised manuscript, we include prior information on ocean fluxes from NASA's ECCO-Darwin product instead of from Takahashi. Recent inverse modeling studies using OCO-2 (e.g., *Liu et al*. 2020) have used the ECCO-Darwin product in place of Takahashi. Existing studies have also shown that ECCO-Darwin exhibits broad consistency with surface ocean pCO_2 observations (e.g., *Carroll et al*., 2020), and the global ocean sink from ECCO-Darwin shows better agreement with the Global Carbon Project (GCP; *Friedlingstein et al*., 2019) than from Takahashi. We have also added extensive additional model-data comparisons using numerous ground-based datasets. These datasets include 20 regular aircraft sites (the Supplemental Sect. S4; Figs. S3-S6), the Atmospheric Tomography Mission (ATom) (Fig. S7), and 18 sites from the Total Carbon Column Observing Network (TCCON) (Figs. S8-S12). We have further evaluated our flux estimate against provisional results from the most recent OCO-2 model-intercomparison (MIP) (shown in Fig. R1 above).

The comparison to other model output was largely limited to the OCO-2 model intercomparison study of Crowell et al. (2019), without following the considerable effort they put into validation or consideration of in-situ measurements. Looking at TCCON sites is an obvious choice, as is the extension to additional aircraft measurements, such as AToM, which are available for at least a couple months of 2016. But comparing your (unclosed) budget to the land biosphere budget of other (mass-conserving) studies is intrinsically misleading. (I am not as surprised that BIC did not pick out the GFED emissions, as these are a few orders of magnitude smaller and are easily swallowed up in the biosphere signal.)

We have included model-data comparisons against both TCCON and ATom in the revised manuscript (the Supplemental Sect. S4 and Figs. S7-S12).

We have also overhauled the inverse modeling setup, and we have reformulated the **X** matrix in the inverse model in a way that ensures the inclusion of more detailed prior information on biomass burning fluxes (We specifically do so by grouping GFED in the same column of **X** with anthropogenic emissions and ocean fluxes.)

L10 & L204-205: While the difference in wording is subtle, I think the abstract overstates what the meteorological variables explain. Do they really describe 90% of the

variability in the fluxes (as seen through OCO-2 observations)? This sort of implies that OCO-2 can "see" fluxes, which isn't true of course. The latter explanation that the deterministic model accounts for XX% of the variance in the estimated fluxes seems more accurate. As you're only treating fluxes on a daily time scale, you're definitely not describing 90% of the variability in the fluxes themselves.

Thank you for this suggestion. We have revised the wording of the manuscript accordingly.

Figure 3 and discussion around L235: This is actually quite interesting! I would be interested in seeing some more analysis of this point. It was also not entirely clear to me what was correlated (and how) in Figure 3. The meteorological variables have been "passed through an atmospheric [transport] model": were they then sampled as column-averaged variables, as OCO-2 views the atmosphere? Were the same averaging kernels applied? It also says that this is the correlation "within different global biomes". Were these columns averaged across space then, and the correlation taken in time? Or is this a spatial correlation coefficient between the column-averaged maps for a given time? I feel like there is an intriguing result here, but I don't fully understand what you've done.

We have added an entire section to the results and discussion to elaborate on this point (Sect. 3.1). We have also revised the analysis described above and instead use synthetic data simulations to better communicate the overarching message of this discussion.

L238 & 239: How can you be sure that this collinearity is playing a bigger role than retrieval or model errors? Would the latter two effects not also limit the model selection?

It is likely that both colinearity and model-data errors play a key role -- in the results of model selection and the uncertainties in the resulting regression coefficients. Numerous existing studies have shown that small biases in space-based CO_2 observations or atmospheric transport models can have significant impacts on CO_2 fluxes estimated using inverse modeling. The purpose of this discussion is not to minimize the importance of these model-data errors, and we have tried to make that clear in the revised manuscript. Rather, we felt that colinearity is an interesting, additional consideration when interpreting the results. The problems caused by colinearity are unrelated to the model-data errors; colinearity depends upon similarities or differences among the predictor variables in the regression, not on the accuracy of the OCO-2 data. Even if the OCO-2 data had perfect accuracy and there were no errors in the GEOS-Chem model, the analysis of relationships between CO_2 fluxes and environmental variables would still be limited by colinearity.

L244 & L260: These statements seem to contradict each other. The first says that the negative beta values for PAR mean that an increase in PAR leads to a decrease in NEE and an increase in uptake. The latter says that the negative beta value for scaled temperature means that an increase in temperature leads to reduced uptake. How can these both be true? This is fundamental to the conclusions drawn.

We have clarified this point in the revised manuscript (Sect. 3.2). An increase in PAR is associated with greater CO_2 uptake by the biosphere (i.e., negative NEE). The scaled temperature function is an upside-down parabola, not a monotonically increasing function. At temperatures below 20 – 25 degrees Celsius, an increase in temperature is associated with negative change in NEE in the inverse model. At temperatures above 20-25 degrees C, an increase in temperature is associated with a positive change in NEE. We also describe this scaled temperature function in detail in the Supplemental Sect. S3 and Fig. S1.

L257-258: While cloudiness is correlated with clouds and rainfall, it's also correlated with the presence or absence of satellite measurements. What impact might this have on your results?

Data sparsity in cloudy regions is certainly an issue for satellite-based greenhouse gas sensors. This issue likely increases the uncertainty in our estimated coefficient for precipitation, particularly in wet climates like tropical forests. It may also be one factor in why we only selected a limited number of environmental driver datasets in many biomes. We point out and discuss this issue in Sect. 3.3 of the revised manuscript.

L302: I'm actually surprised Australia matches as well as it does, as you've had to fold the Southern Ocean sink into the Southern Hemisphere land fluxes somehow.

We did not do a good job of describing the treatment of ocean fluxes in the inverse model. We have both improved the description of ocean fluxes and have overhauled the inverse modeling setup to more explicitly include a prior ocean flux estimate within the inverse model.

RC3: Abhishek Chatterjee

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.

We have re-written the manuscript and focused on the second question described above (the connections between CO_2 fluxes and environmental drivers). We have also de-emphasized the technical or methodological components. We hope that the revised manuscript has a much clearer formulation in terms of 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).

We have greatly expanded model evaluation and have overhauled the inverse modeling setup, including the model-data-mismatch variance. These points are discussed in greater detail below in reply to individual reviewer comments.

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.

We have re-written the manuscript to focus on the relationships between carbon fluxes and environmental drivers and have de-emphasized the inversion framework or methodology. We feel that these environmental relationships make for a more interesting scientific study than focusing on methodological questions, and we hope that this re-write has yielded a manuscript with a much clearer purpose and scope. As part of this revision, we have expanded the time period of the study from one year (2016) to four years (2015 - 2018). In addition, we have included extensive comparisons with terrestrial biosphere models (TBMs) to improve the depth and novelty of the analysis in the manuscript (Sect. 3.3). Specifically, in the revised manuscript, we compare the environmental relationships that we infer from OCO-2 with the environmental relationships that we infer from 15 state-of-the-art TBMs from the recent TRENDY model comparison project.

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 midto-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.

We have added substantial new analysis to the revised manuscript to improve the novelty and depth of the scientific results. Specifically, we not only infer environmental relationships using observations from OCO-2 but also compare those against the environmental relationships inferred from 15 TBMs for the same time period. Using OCO-2, we find stronger relationships between temperature and CO_2 fluxes across tropical biomes compared to many TBMs, and we find that increases in precipitation across the tropics are associated with greater carbon uptake across seasonal time scales and biome-level spatial scales, a result that disagrees with about half of the TBMs that estimate the opposite relationship. Overall, there are large uncertainties in the environmental relationships within TBMs across all global biomes. The relationships with precipitation are most uncertain in these models while TBMs show greatest agreement on the relationships with temperature. This disagreement over the relationship with precipitation may be due, at least in part, to large disagreements over the fate of precipitation in these ecosystems; each the TRENDY models input the same precipitation estimate but yield evapotranspiration that differs by up to a factor of three among models, depending upon the season and biome. The revised manuscript highlights both the opportunities for informing TBM development using atmospheric observations but also the challenges of doing so using current satellite-based datasets of CO_2.

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?

We have greatly expanded the discussion of the auxiliary variables in the re-written manuscript. For example, we have included scatter plots showing the estimated coefficients for each year (Fig. 4), compared those coefficients against coefficients estimated from 15 TBMs (Fig. 3a), and showed the coefficient of variation (as suggested by the reviewer, Fig. 3b). All of the coefficients in the manuscript are listed in units of flux (μmol m^{-2} s^{-1}), so we can better compare the coefficients among different auxiliary variables and different biomes.

Note that in the revised manuscript, we have included PAR instead of fPAR. This was an oversight on our behalf. Furthermore, we decided not to include remote sensing indices in this manuscript because we wanted to focus on comparing the environmental processes in state-of-the-art TBMs against the relationships that we infer from OCO-2. Some TBMs use remote sensing indices like SIF, but some do not. Hence, we felt that it was more appropriate to focus on environmental processes instead of vegetation indices like SIF or LAI that may not be applicable to many of the TBMs compared in the manuscript. Hence, all of the auxiliary variables used in the revised manuscript are from meteorological reanalysis. We wanted to clearly focus the scope of this manuscript on environmental processes, but we think that an examination of remote sensing indices and global carbon fluxes would make for an interesting future study.

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.

We have greatly expanded model-data comparisons in the revised manuscript. In the new manuscript, we evaluate the model-data residuals both for the full posterior flux estimate and for the component of the fluxes that is described by the auxiliary variables. In addition, we compare against numerous independent datasets, including 20 regular aircraft sites (the Supplemental Sect. S4; Figs. S3-S6), the Atmospheric Tomography Mission (ATom) (Fig. S7), and 18 sites from the Total Carbon Column Observing Network (TCCON) (Figs. S8-S12). We also provide model-data evaluations for each year of the four-year study period to show that there is no trend in the model-data comparisons (Fig. S2).

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.

Several of the reviewers, including this reviewer, recommended defining a more targeted scope and more clearly defined aims in the manuscript, and we have tried to do so in the re-written manuscript. The focus of this manuscript is estimating the relationships between CO_2 fluxes and environmental driver datasets using OCO-2 and comparing those inferences against the relationships estimated from 15 state-of-the-art TBMs. We agree that an *in situ* data study would

be interesting, but we feel that this focus would be better left for a separate study in the interest of maintaining a targeted scope with clearly defined aims. Furthermore, the results and discussion section of the revised manuscript are heavily focused on the tropics, and the in-situ observation network is very sparse across the tropics; existing studies have raised questions about the strength of the tropical flux constraint in in-situ inversions (e.g., *Crowell et al*. 2019; *Piao et al*. 2020).

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 ppm2 based on a previous pseudo-data study. Why? I highly encourage the authors to use the reported XCO_2 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).

We overhauled the inverse modeling setup in response to suggestions from reviewers and have changed the covariance matrix parameters in the inverse model as suggested by this reviewer. Specifically, we use estimated model-data mismatch errors from the 10-second OCO-2 data product (e.g., *Crowell et al*. 2019), described in Sect. 2.4 and the Supplemental Sect. S1. In addition, we use an exponential model for the **Q** covariance matrix. Note that a spherical is very similar to an exponential model, but a spherical model decays to zero, unlike an exponential model which decays to near-zero but never actually reaches zero (e.g., *Kitanidis*, 1997). A spherical model therefore yields covariance matrices that require substantially less computer memory, a particular benefit for large inverse problems (e.g., *Miller et al*. 2020). In this study, the components of **Q** are small enough such that we were able to use an exponential model.

References:

Carroll, D., Menemenlis, D., Adkins, J.F., Bowman, K.W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M.M., Hill, C., Jahn, O. and Landschützer, P.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean pCO_2 and Air-Sea CO_2 Flux. Journal of Advances in Modeling Earth Systems, *12*(10), p.e2019MS001888. https://doi.org/10.1029/2019MS001888, 2020.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmos. Chem. Phys., 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.

Fang, Y., and Michalak, A. M.: Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers. Global Biogeochemical Cycles, 29(5), 555-566. https://doi.org/10.1002/2014/GB005034, 2015.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth Syst. Sci. Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Huntzinger, D. N., Gourdji, S. M., Mueller, K. L., and Michalak, A. M.: A systematic approach for comparing modeled biospheric carbon fluxes across regional scales, Biogeosciences, 8, 1579–1593, https://doi.org/10.5194/bg-8-1579-2011, 2011.

Kitanidis, P.: Introduction to Geostatistics: Applications in Hydrogeology, Stanford-Cambridge program, Cambridge University Press, Cambridge, 1997.

Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, N. C., Oda, T., Carroll, D., Menemenlis, D., Joiner, J., Commane, R., Daube, B., Gatii, L. V., McKain, K., Miller, J., Stephens, B. B., Sweeney, C., and Wofsy, S.: Carbon Monitoring System Flux Net Biosphere Exchange 2020 (CMS-Flux NBE 2020), Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-123, in review, 2020.

Miller, S. M., Michalak, A. M., Yadav, V., and Tadić, J. M.: Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite, Atmos. Chem. Phys., 18, 6785–6799, https://doi.org/10.5194/acp-18-6785-2018, 2018.

Miller, S. M. and Michalak, A. M.: The impact of improved satellite retrievals on estimates of biospheric carbon balance, Atmos. Chem. Phys., 20, 323–331, https://doi.org/10.5194/acp-20-323-2020, 2020.

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., ... & Sitch, S.: Interannual variation of terrestrial carbon cycle: Issues and perspectives. Global Change Biology, 26(1), 300-318, 2020.

Linking global terrestrial CO_2 fluxes and environmental driversusing ~~OCO-2~~ : inferences from the Orbiting Carbon Observatory-2 and ~~a geostatistical inverse model~~terrestrial biospheric models

Zichong Chen[1], Junjie Liu[2], Daven K. Henze[3], Deborah N. Huntzinger[4], Kelley C. Wells[5], Stephen Sitch[6], Pierre Friedlingstein[7], Emilie Joetzjer[8], Vladislav Bastrikov[9], Daniel S. Goll[10], Vanessa Haverd[11], Atul K. Jain[12], Etsushi Kato[13], Sebastian Lienert[14], Danica L. Lombardozzi[15], Patrick C. McGuire[16], Joe R. Melton[17], Julia E. M. S. Nabel[18], Benjamin Poulter[19], Hanqin Tian[20], Andrew J. Wiltshire[21], Sönke Zaehle[22], and Scot M. Miller[1]

[1]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[3]Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO, USA
[4]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ, USA
[5]Department of Soil, Water, and Climate, University of Minnesota-Twin Cities, St. Paul, MN, USA
[6]College of Life and Environmental Sciences, University of Exeter, Exeter, UK
[7]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
[8]Centre National de Recherche Meteorologique, Unite mixte de recherche 3589 Meteo-France/CNRS, 42 Avenue Gaspard Coriolis, 31100 Toulouse, France
[9]Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, CEA-CNRS-UVSQ, CE Orme des Merisiers, 91191 Gif-sur-Yvette CEDEX, France
[10]Université Paris Saclay, CEA-CNRS-UVSQ, LSCE/IPSL, Gif sur Yvette, France
[11]CSIRO Oceans and Atmosphere, G.P.O. Box 1700, Canberra, ACT 2601, Australia
[12]Department of Atmospheric Sciences, University of Illinois, Urbana, IL, USA
[13]Institute of Applied Energy (IAE), Minato-ku, Tokyo 105-0003, Japan
[14]Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[15]National Center for Atmospheric Research, Climate and Global Dynamics, Terrestrial Sciences Section, Boulder, CO, USA
[16]Department of Meteorology, Department of Geography & Environmental Science, National Centre for Atmospheric Science, University of Reading, Reading, UK
[17]Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada
[18]Max Planck Institute for Meteorology, Hamburg, Germany
[19]NASA Goddard Space Flight Center, Biospheric Sciences Laboratory, Greenbelt, MD, USA
[20]International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, 602 Ducan Drive, Auburn, AL, USA
[21]Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK
[22]Max Planck Institute for Biogeochemistry, P.O. Box 600164, Hans-Knöll-Str. 10, 07745 Jena, Germany

Correspondence: Zichong Chen (zchen74@jhu.edu)

Abstract. Observations from the Orbiting Carbon Observatory 2 (OCO-2) satellite ~~, launched in July 2014,~~ have been used to estimate CO_2 fluxes in many regions of the globe and provide new insight ~~on~~ into the global carbon cycle. ~~A challenge now is to not only estimate fluxes using satellite observations but also to understand how these fluxes are connected to variations in environmental conditions. In this study , we specifically evaluate the capabilities and limitations of utilizing current~~ The objective of this study is to infer the relationships between patterns in OCO-2 ~~observations to infer connections between CO_2 fluxes and underlying environmental variables.To do so, we adapt geostatistical inverse modeling to satellite-based applications and evaluate a case study for year 2016~~ observations and environmental drivers (e.g., temperature, precipitation) and therefore inform a process understanding of carbon fluxes using OCO-2. ~~A unique aspect of the geostatistical approach is that we can use estimates of environmental and meteorological variables to help estimate CO_2 fluxes in place of a traditional prior flux model . We~~ We use a multiple regression and inverse model, and the regression coefficients quantify the relationships between observations from OCO-2 and environmental driver datasets within individual years for 2015–2018 and within seven global biomes. We subsequently compare these inferences to the relationships estimated from 15 terrestrial biosphere models (TBMs) that participated in the TRENDY model inter-comparison. Using OCO-2, we are able to quantify ~~the relationships between~~ only a limited number of relationships between patterns in atmospheric CO_2 ~~fluxes and a few environmental variables across global biomes; we find that a simple combination of air temperature, daily precipitation, and photosynthetically active~~ observations and patterns in environmental driver datasets (i.e., 10 out of the 42 relationships examined). We further find that the ensemble of TBMs exhibits a large spread in the relationships with these key environmental driver datasets. The largest uncertainty in the models is in the relationship with precipitation, particularly in the tropics, with smaller uncertainties for temperature and photosynthetically-active radiation (PAR)~~can describe almost 90% of the variability in CO~~. Using observations from OCO-2, we find that precipitation is associated with increased CO_2 ~~fluxes as seen through~~ uptake in all tropical biomes, a result that agrees with half of the TBMs. By contrast, the relationships that we infer from OCO-2 ~~observations. PAR is an adept predictor of fluxes across mid-to-high latitudes, whereas a combined set of air temperature and precipitation shows strong explanatory power across tropical biomes~~for temperature and PAR are similar to the ensemble mean of the TBMs, though the results differ from many individual TBMs. These results point to the limitations of current space-based observations for inferring environmental relationships but also indicate the potential to help inform key relationships that are very uncertain in state-of-the-art TBMs. ~~However, we are unable to quantify relationships with additional environmental variables because many variables are correlated or colinear when passed through an atmospheric model and averaged across a total atmospheric column. Overall, we estimate a global net biospheric flux of -1.73 ± 0.53 GtC in year 2016, in close agreement with recent inverse modeling studies using OCO-2 retrievals as observational constraints.~~

1 Introduction

Over the past decade, the field of space-based CO_2 monitoring has undergone a rapid evolution. The sheer number of CO_2-observing satellites has greatly increased, including GOSAT/GOSAT-2 (Kuze et al., 2009; Nakajima et al., 2012), TanSat (Yang et al., 2018) and OCO-2/OCO-3 (Crisp, 2015; Eldering et al., 2019). ~~These dramatically expanded satellites~~

observe This expanding observing system provides atmospheric CO_2 observations broadly across the globe, making it possible to estimate the distribution and magnitude of CO_2 fluxes in many regions of the globe that previously had sparse *in situ* that have sparse in situ surface atmospheric CO_2 monitoring (e.g., the tropics and the Southern Hemisphere). For example, the OCO-2 satellite, launched in July 2014, provides ~65,000 high-quality observations per day (Eldering et al., 2017); the that pass quality screening (Eldering et al., 2017); this dense, global set of OCO-2 observations, combined with inverse modeling techniques, have has been used to constrain regional- and continental-scale CO_2 sources and sinks and provide new insights into CO_2 fluxes (e.g., Liu et al., 2017; Crowell et al., 2019; Palmer et al., 2019) (e.g., Eldering et al., 2017; Liu et al., 2017; Crowell et al., 2019; Palmer et al., 2019; Byrne et al., 2020a).

Furthermore, recent Recent advances in OCO-2 retrievals from the NASA ACOS science team have led to widespread improvements in the observations (e.g., O'Dell et al., 2018), and these improvements have enabled increasingly accurate and detailed CO_2 flux constraints from inverse modeling (Miller and Michalak, 2020)reductions in observation errors (e.g., O'Dell et al., 2018). Reducing the biases in satellite retrievals errors in satellite observations of CO_2 is critical for understanding CO_2 sources and sinks using inverse modeling, as even small retrieval biases can have a large biases in the observations can have an impact on the CO_2 flux estimate (e.g., Chevallier et al., 2014; Miller et al., 2018) (e.g., Chevallier et al., 2007; Feng et al., 2016; Chevallier et al., 2014; Miller et al., 2018). For example, Miller et al. (2018) evaluated the extent to which OCO-2 retrievals can detect patterns in biospheric CO_2 fluxes and found that an early version of the OCO-2 retrievals (version 7) is only equipped to provide accurate flux constraints across very large continental or hemispheric regions; by contrast, in a companion follow-up paper, Miller and Michalak (2020) re-visited satellite capabilities in light of recently improved OCO-2 retrievals, and the authors suggested argued that new OCO-2 retrievals can be used to constrain CO_2 fluxes for more detailed regions (i.e., for seven global biomes).

A challenge now is to not only estimate the magnitude further challenge is to use these new global satellite datasets to evaluate and improve process-based estimates of the global carbon cycle provided by terrestrial biospheric models (TBMs). TBMs have become an integral tool for understanding regional- and global-scale carbon dynamics and for predicting future carbon cycling under changing climate. With that said, existing TBMs show large uncertainties in carbon flux estimates at multiple spatial and temporal scales – at regional and seasonal scales (e.g., Peng et al., 2014; King et al., 2015), at global and inter-annual scales (e.g., Piao et al., 2020), and in historical and future projections (e.g., Friedlingstein et al., 2006; Huntzinger et al., 2017).

One approach to inform TBM development is to estimate flux totals using atmospheric observations and compare those totals against TBMs – to inform the magnitude, seasonality, or spatial distribution of fluxes using these new OCO-2 retrievals but also to understand how variations in fluxes are connected to variations in environmental drivers (e.g., King et al., 2015; Bastos et al., 2018). A more challenging approach is to estimate the relationships between CO_2 fluxes and environmental drivers using atmospheric observations and compare those relationships directly to the relationships in TBMs. We define the term "environmental drivers" as any meteorological variables or characteristics of the physical environment that can be modeled or measured and may correlate with net ecosystem exchange (NEE). Existing studies on the capability of satellite observations have widely focused on constraining the magnitude and distribution of fluxes

(Eldering et al. , 2017; Liu et al., 2017; Palmer et al., 2019; Crowell et al., 2019; Chevallier et al., 2019). It is now time to push these satellite observations further and explore whether the observations can be used to infer connections between fluxes Several studies have shown that these types of comparisons are feasible using in situ atmospheric observations (e.g., Dargaville et al., 2002; Forkel et al., 2016; Gourdji et al., 2008, 2012; Piao et al., 2013, 2017; Wang et al., 2014; Fang and Michalak . Among other studies, Fang and Michalak (2015) used in situ atmospheric CO_2 observations across North America and an inverse modeling framework to probe the relationships between NEE and environmental drivers(refer to hereafter as 'connections') across many different regions of the globe. Variations in CO; the authors compared these relationships directly to those inferred from several TBMs, and found that TBMs have reasonable skill in representing the relationship with shortwave radiation but show weak performance in describing relationships with other drivers like water availability. Similarly, Shiga et al. (2018) used tower-based atmospheric CO_2 observations to explore regional interannual variability (IAV) in NEE across North America, and found that TBMs disagree on the dominant regions responsible for IAV; this disagreement can be linked to differing sensitivities of CO_2 fluxes are closely linked with variations in environmental drivers , and understanding these connections is key if we are to use these new satellite observations to evaluate and improve process-based terrestrial biospheric models (TBMs) .

These connections have been extensively studied at local and global scales. At site levels (1 km^2), eddy covariance flux tower measurements have provided excellent detail to quantify these connections (e. g., Desai et al., 2010; Baldocchi et al., 2017); At the global level, existing studies (e.g, Wang et al., 2014; Keppel-Aleks et al., 2014; Piao et al., 2013) used fluxes to environmental drivers within the TBMs. At even longer temporal scale, Wang et al. (2014) employed atmospheric CO_2 from global background stations (e.g., growth rate record from Mauna Loa, Hawaii, USA and the South Pole) to illustrate a global picture of these connections. However, intermediate, regional-scale connections are still poorly understood (e.g., Niu et al., 2017; Shiga et al., 2018). To date, previous studies have used ground-based and aircraft observations of atmospheric for five decades to explore the sensitivity of the global CO_2 to link the fluxes and underlying environmental processes (e.g., Gourdji et al., 2012; Fang and Michalak, 2015; Fang et al., 2017; Shiga et al., 2018; Hu et al., 2019). However, it is difficult to constrain these connections for regions with few *in situ* atmospheric COgrowth rate to tropical temperature; the authors found that existing TBMs do not capture the observed sensitivity of the growth rate to tropical climatic variability, implying a limited ability of these TBMs in representing the impact of drought and warming on tropical carbon dynamics.

More recently, a handful of studies have shown that it is possible to tease out relationships between CO_2 observations (e.g., the tropics and the Southern Hemisphere). The global coverage of fluxes and environmental drivers using global satellite observations of CO_2 (e.g., Liu et al., 2017; Byrne et al., 2020b). For example, Liu et al. (2017) used observations from OCO-2 observations provides a novel opportunity to bridge the gap and explore these connections on region scales (e. g., Eldering et al., 2017; Liu et al. 2017).

However, it is still unclear the extent to which we can make these regional-scale connections given the accuracy and coverage of current OCO-2 observations . Indeed, Liu et al. (2017) used net biosphere fluxes inferred from version seven OCO-2 retrievals along with component carbon fluxes to disentangle to disentangle the environmental processes related to the flux anomalies in tropical regions during the 2015-2016 El Niño. However, Chevallier (2018) suggested that the satellite retrievals

used in Liu et al (2017) cannot provide sufficient accuracy and sensitivity to separately constrain continental flux anomalies and associated environmental processes over the tropics. Hence, we specifically evaluate the capability and limitation of using current Byrne et al. (2020b) assimilated in situ and GOSAT observations of atmospheric CO_2 and an inverse model framework, and found contrasting environmental sensitivities of IAV in CO_2 fluxes between western and eastern temperate North America.

The goal of this study is to use atmospheric CO_2 observations from OCO-2 retrievals to infer these connections on regional scales using a geostatistical inverse modeling to quantify the relationships between spatiotemporal patterns in CO_2 fluxes and patterns in environmental driver datasets. We conduct this analysis for years 2015 – 2018 and focus on relationships that manifest across an individual year and individual biome. We specifically quantify the relationships using a top-down regression framework and a geostatistical inverse model (GIM).

A GIM is particularly well-suited to systematically evaluating these connections. Specifically, a GIM does not prescribe or rely on a traditional prior flux model . The choice of prior fluxes in a classical inverse model is often subjective, and this choice can impact the posterior flux estimate (e.g., Peylin et al., 2013; Houweling et al., 2015; Philip et al., 2019) . By contrast, a GIM can assimilate a wide range of environmental drivers, making it possible to evaluate data-driven connections between these variations in environmental drivers and CO_2 fluxes inferred from atmospheric observations (see Sect. We then compare the relationships inferred using OCO-2 observations against those inferred from 15 state-of-the-art TBMs from the TRENDY model comparison project (v8, https://sites.exeter.ac.uk/trendy; see Table S1 for a full list of TBMs; Sitch et al., 2015; Friedlingstein et al., 2019) . The primary objectives of this analysis are threefold: (1) evaluate what kinds of environmental relationships we can infer using current satellite observations from OCO-2, (2) . Existing GIM studies have investigated connections of assess where and when TBMs do and do not show consensus on the relationships between CO_2 fluxes and environmental drivers for North America (Gourdji et al., 2010salient environmental drivers, 2012; Commane et al., 2017; Shiga et al. 2018) and the globe (Gourdji et al., 2008) using a variety of *in situ* CO_2 observations.

GIMs, however, have never been applied to global satellite observations, and the extension of GIMs from small, regional *in situ* datasets to a massive, global satellite datasets like and (3) compare the relationships inferred from OCO-2 presents novel computational and statistical challenges. To overcome this challenge, we combine the GIM with the adjoint of a global chemical transport model. Using this framework, we not only estimate daily global COagainst those inferred from TBMs with the goal of informing and improving TBM development.

2 Methods

2.1 Overview

We quantify the relationships between CO_2 fluxes at the model grid scale (latitude × longitude) but also quantify posterior uncertainties in the estimated fluxes. This study builds upon previous efforts (Miller et al., 2018; Miller and Michalak, 2020) in which the authors evaluated when and where the OCO-2 observations can be used to constrain biospheric CO_2 fluxes. In this

study we push one step further and explore the connections observations from OCO-2 and environmental driver datasets for different regions of the globe using a top-down regression framework and a GIM. We cannot directly observe the relationships between CO_2 fluxes and environmental drivers. The primary purpose driver datasets. With that said, an overarching idea of this study is to couple a GIM to a global adjoint model and use this framework to systematically evaluate what kind of regional-scale connections we can (and cannot) make using current OCO-2 observations . We focus on a single year (i.e., 2016) as an initial case study – to explore the applicability of the geostatistical approach to large satellite-based inverse problems. We first describe the implementation of the GIM for that these relationships manifest in atmospheric CO_2 observations, and we can quantify at least some of these relationships using observations from OCO-2 observations; we then evaluate and discuss the results of this approach using the 2016 exploratory case studyand a weighted, multiple regression. The coefficients estimated as part of the regression relate patterns in atmospheric CO_2 observations to patterns in the environmental driver datasets.

2. Data and Methods

2.1 Approach overview

We design a framework that couples the GIM to a global adjoint model (version v35n of the GEOS-Chem adjoint, Henze et al., 2007) and explore the applicability of the geostatistical approach to inverse problems with a large number of flux grid boxes As part of this analysis, we also explore differences in the estimated environmental relationships (i.e., 1.2×10^6) and a large number of OCO-2 satellite observations (i. e., 9×10^4). We use year 2016 as an initial case study, as there is better temporal coverage of good-quality data from OCO-2 throughout the entire year relative to years regression coefficients) among different years and different biomes. To this end, we estimate separate regression coefficients for each of seven different global biomes, and we estimate separate coefficients for each individual year of the study period (2015 – 2018). Hence, each coefficient estimated here represents the relationship between OCO-2 observations and an environmental driver dataset across an entire year and 2017. For example, there are 7 week-long gaps in the a global biome. Miller and Michalak (2020) explored when and where current OCO-2 data in year 2015 and a 1.5-month gap in the OCO-2 data in year 2017, whereas there are no such gaps in year 2016. This time period also overlaps with an OCO-2 inverse modeling inter-comparison (MIP) study, enabling direct comparison with those results (Crowell et al., 2019). We specifically estimate observations can be use to detect variability in surface CO_2 fluxesfor September 1, 2015 to December 31, 2016 but discard the first four months as a spin-up time period. We also offer up a wide range of environmental drivers and allow the GIM to select a subset thatbest predicts spatiotemporal patterns in CO_2 fluxes at the model grid scale, described in detail below (Sects. 2.2-2.4and the authors argue that, in most seasons, the satellite can be used to constrain fluxes from seven large biome-based regions. Hence the choice of the seven biomes used in this study (Fig. 1).

2.2 OCO-2 satellite observations

We utilize 10-s average XCO2 generated from version 9 of the satellite observations for the period from September 1, 2015 through the end of year 2016 (e.g., Chevallier et al., 2019) . We use both land nadir- and land glint-mode retrievals in the inverse model . Recent retrieval updates have eliminated biases that previously existed between land nadir and land glint observations (O'Dell et al., 2018We first conduct this analysis using CO_2 observations from OCO-2. We then conduct a parallel analysis using the outputs of 15 terrestrial biosphere models (TBMs) from the TRENDY model inter-comparison

project (v8). ~~Moreover, Miller and Michalak (2020) evaluated the impact of these updated~~ The goal of this step is to compare the environmental relationships (i.e., regression coefficients) that we infer from OCO-2 ~~retrievals on the terrestrial CO_2 flux constraint in different regions of the globe; the authors found that~~ against the regression coefficients that we estimate from numerous state-of-the-art TBMs. We can then identify any similarities or differences between the TBMs and inferences using OCO-2 observations. We specifically analyze TRENDY model outputs for years 2015–2018, the same years as the ~~inclusion of both land nadir and land glint retrievals yielded a stronger constraint on CO_2 fluxes relative to using only a single observation type.~~

~~2.3 Geostatistical inverse model~~

~~A GIM does not require an emission inventory or a bottom-up model as an initial guess of fluxes; instead, a GIM can leverage a wide range of environmental driver datasets to help predict spatial and temporal patterns in the CO_2 fluxes (e. g., Gourdji et al., 2008, 2012; Shiga et al., 2018). We further pair the GIM with a statistical approach known as model selection to objectively determine which set of drivers can best reproduce CO_2 observationsfrom~~ OCO-2 analysis described above. To conduct this analysis, we generate synthetic OCO-2 observations using each of the 15 TBMs and using an atmospheric transport model. We then run the multiple regression on these synthetic observations. This setup ~~makes it feasible to both estimate CO_2 fluxes and to explicitly quantify the relationships between the fluxes and the underlying environmental drivers. The fluxes, as estimated by the GIM, consist of two components. First, the GIM will scale the environmental drivers to match patterns in the atmospheric observations , and this component of the flux estimate is referred to as the 'deterministic model'. Second, the GIM will model space-time patterns in the CO_2 fluxes that are implied by the atmospheric observations but not explained by any environmental drivers, and this component of the fluxes is referred to as the 'stochastic component'. The best flux estimate is a sum of the deterministic model and the stochastic component:~~ mirrors that of Fang and Michalak (2015) and creates an apples-to-apples comparison between the TBMs and OCO-2 observations; in each case, we use atmospheric observations (either real or synthetic) and use the same set of equations to estimate the regression coefficients.

The multiple regression used in this study has the following mathematical form (e.g., Fang and Michalak, 2015):

$$\boldsymbol{z} = h(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\zeta}) + \boldsymbol{\epsilon} \tag{1}$$

where ~~$\boldsymbol{s}$ are $m \times 1$ unknown fluxes, $\mathbf{X}$ is a~~ $\boldsymbol{z}$ $(n \times 1)$ is a vector of real or synthetic CO_2 observations from OCO-2, $\mathbf{X}$ $(m \times p)$ is a matrix of environmental ~~drivers (see Sect. 2.4),~~ driver datasets (described in Sect. 2.2), and $\boldsymbol{\beta}$ ~~are~~ $(p \times 1$ ~~unknown scaling factors or drift coefficients. These coefficients quantify the relationships between each of the p environmental drivers (i. e., each~~) are the regression coefficients that are estimated as part of the regression. Each column of $\mathbf{X}$ ~~) and the estimated CO_2 fluxes. The product of $\mathbf{X}$ and $\boldsymbol{\beta}$ is the deterministic model($\mathbf{X}\boldsymbol{\beta}$). The stochastic component (~~represents a different environmental driver dataset for a specific biome in a specific year. Note that we estimate all of the coefficients for the different environmental drivers and different biomes simultaneously in the regression model. In addition, $\boldsymbol{\zeta}$ ~~) is zero-mean with a pre-specified spatial and/or temporal correlation structure; it describes spatial and temporal~~ $(m \times 1)$ represents patterns in the fluxes that ~~are not captured by the deterministic model. For the setup here, the drift coefficient ($\boldsymbol{\beta}$) associated with each environmental driver is constant in space and time, while the stochastic component~~ cannot be described by the environmental driver datasets, and these values are

unknown. This component of the fluxes is also commonly referred to as the stochastic component and is discussed in Sect. 2.5. $h()$ is an atmospheric transport model ($\boldsymbol{\zeta}$) ~~varies at the model grid scale.~~ described later in this section) that relates surface CO_2 fluxes ($\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\zeta}$) to the atmospheric CO_2 observations, and $\boldsymbol{\epsilon}$ ($n \times 1$) is a vector of errors in the OCO-2 observations and/or in the atmospheric model. The statistical properties of these errors are estimated before running the regression (described in Sect. 2.4).

~~We estimate both the fluxes ($\boldsymbol{s}$) and the drift coefficients ($\boldsymbol{\beta}$) by minimizing the GIM cost function (e.g., Kitanidis and Vomvoris, 1983; Kitanidis, 1995; Michalak et al. , 2004):~~

$$L_{\boldsymbol{s},\boldsymbol{\beta}} = \tfrac{1}{2}(\boldsymbol{z} - h(\boldsymbol{s}))^T \mathbf{R}^{-1}(\boldsymbol{z} - h(\boldsymbol{s})) + \tfrac{1}{2}(\boldsymbol{s} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}^{-1}(\boldsymbol{s} - \mathbf{X}\boldsymbol{\beta})$$

~~The cost function includes two components: the first component indicates that the fluxes ($\boldsymbol{s}$), when run through an atmospheric model, $h(\boldsymbol{s})$, should match the observations ($\boldsymbol{z}$) to within a specific error tolerance ($\boldsymbol{z} - h(\boldsymbol{s})$) that is prescribed by the covariance matrix $\mathbf{R}$ ($n \times n$) . $\mathbf{R}$ describes model-data mismatch errors , including errors from the atmospheric transport model and~~Note that this framework assumes linear relationships between the environmental driver datasets and the OCO-2 ~~retrievals, among other errors. The second component of Eq. 2 stipulates that the structure of the stochastic component ($\boldsymbol{s} - \mathbf{X}\boldsymbol{\beta}$) is described by the covariance matrix $\mathbf{Q}$ ($m \times m$). $\mathbf{Q}$, like $\mathbf{R}$, must be defined by the modeler before estimating the fluxes; it represents the variances and spatiotemporal covariances of the stochastic component. We estimate $\mathbf{Q}$ using a statistical approach known as Restricted Maximum Likelihood (RML; e.g., Kitanidis, 1997; Gourdji et al., 2012; Miller et al., 2016). $\mathbf{Q}$ includes both diagonal and off-diagonal elements; the latter decay with the separation time and distance between two model grid boxes. We construct $\mathbf{R}$ as a diagonal matrix with constant elements on the diagonal. The Supplement Sect. S1 provides a detailed explanation of the approach used here to estimate the covariance matrix parameters.~~

~~After estimating the covariance matrix parameters, we then estimate the~~ observations. Numerous existing studies have used linear models to approximate relationships with environmental driver datasets. For example, studies have used linear models to compare the relationships between CO_2 fluxes ~~by iteratively minimizing Eq. 2 using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS, Liu and Nocedal, 1989). We use this approach to simultaneously estimate both $\boldsymbol{s}$ and $\boldsymbol{\beta}$. Miller et al (2019)describe this iterative approach to minimize Eq. 2 in detail.~~ and environmental driver datasets in TBMs (e.g., Huntzinger et al., 2011) , to infer these relationships using eddy flux observations (e.g., Mueller et al., 2010; Yadav et al., 2010) , and to infer relationships between atmospheric CO_2 observations and environmental driver datasets (e.g., Gourdji et al., 2012; Fang et al., 2014; Fang and Michalak, 2015; Piao et al., 2013, 2017; Rödenbeck et al., 2018).

~~2.4 Auxiliary environmental drivers~~

~~We consider a wide range of environmental drivers ($\mathbf{X}$).These are meteorological variables primarily related to heat, water, and radiation, available from NASA's~~ The equations above require an atmospheric transport model ($h()$). We use the forward GEOS-Chem model (version v9-02; http://www.geos-chem.org) in this study, and we further use wind fields from the Modern-Era Retrospective Analysis for Research and Applications ~~, Version 2~~ (MERRA-2~~, Rienecker et al. , 2011)~~) to drive

atmospheric transport within GEOS-Chem (Gelaro et al., 2017). The GEOS-Chem simulations used here have a global spatial resolution of 4° latitude by 5° longitude and therefore are best able to capture broad, regional spatial patterns in atmospheric CO_2. ~~Specifically, we consider daily 2-m~~

2.2 Environmental driver datasets

We estimate the relationships between OCO-2 observations (either real or synthetic) and environmental driver datasets drawn from commonly-used meteorological reanalysis. We specifically consider the following driver datasets as predictor variables in the multiple regression: 2-meter air temperature, ~~daily precipitation, 30-day average precipitation,~~ precipitation, photosynthetically active radiation (PAR), ~~surface~~ downwelling shortwave radiation, ~~soil temperature at 10-cm depth, soil moisture at 10-cm depth, specific humidity, and relative humidity.~~ and specific humidity.

We also include a non-linear function of ~~2-m~~ 2-meter air temperature as an environmental driver ~~(refer~~ dataset in the regression (refered to hereafter as scaled temperature~~). This function~~ ; plotted in Fig. S1 and described in detail in Supplement Sect. S3). Numerous existing studies show that the relationship between temperature and photosynthesis has a different sign depending upon the temperature range; at sufficiently warm temperatures, an increase in temperature yields a decrease in photosynthesis (e.g., Baldocchi et al., 2017). The scaled temperature function considered here can account for those differences, and we find that this function yields a better model-data fit in the regression analysis than using temperature alone. The scaled temperature function used here is from the Vegetation Photosynthesis and Respiration Model (VPRM~~, Mahadevan et al., 2008)~~) (Mahadevan et al., 2008) and describes the non-linear relationship between temperature and photosynthesis ~~(e.g. , Raich et al. 1991, see the Supplement Sect. S2).~~ (Raich et al., 1991). The function is shaped like an upside-down parabola (shown in Fig. S1). Furthermore, this type of non-linear temperature function has been commonly used in existing TBMs (e.g., Heskel et al., 2016; Luus et al., 2017; Dayalu et al., 2018; Chen et al., 2019).

The environmental driver datasets described above are drawn from the Climatic Research Unit (CRU) and Japanese Reanalysis (JRA) meteorology product (CRUJRA; Harris, 2019). We use environmental driver data from CRUJRA because it is the same product used to generate the TRENDY model estimates. All flux outputs from TRENDY are provided at a monthly temporal resolution, so we input monthly meteorological variables from CRUJRA into the regression framework. Furthermore, we regrid the environmental driver datasets to a 4° latitude by 5° longitude spatial resolution before inputting these datasets into the regression. This spatial resolution matches the resolution of the atmospheric transport simulations used in this study (described in Sect. 2.1). The regression coefficients therefore quantify the relationships between OCO-2 observations and patterns in environmental driver datasets that manifest at this spatial and temporal resolution.

We subsequently re-run the regression analysis using environmental driver datasets drawn from a second meteorological product. Estimates of environmental driver data like temperature or precipitation can vary among meteorological models, and these differences among models are a source of uncertainty in the estimated regression coefficients. Hence, the use of a second meteorological product can at least partially account for these uncertainties. We specifically re-run the regression analysis using environmental driver datasets drawn from MERRA-2. We choose MERRA-2 because it is a commonly used, global reanalysis product from the NASA Global Modeling and Assimilation Office (GMAO). Furthermore, we use wind fields from MERRA-2

to drive all atmospheric transport model simulations in this study (described in Sect. 2.1), so the use of MERRA-2 for the environmental driver datasets in the regression creates consistency with the wind fields in the atmospheric model simulations that support the regression.

Note that we do not include any remote sensing indices (e.g., ~~solar-induced chlorophyll fluorescence (SIF)~~ solar-induced chlorophyll fluorescence or leaf area index~~(LAI)~~) in the present study. Rather, the focus of this study is to explore environmental drivers of CO_2 fluxes, not remote sensing proxies for CO_2 fluxes.

~~We group the globe into seven biome-based regions and allow the GIM to use different environmental drivers in different biomes. Miller and Michalak (2020) found that current OCO-2 retrievals can be used to constrain terrestrial CO_2 fluxes for regions of this size. The seven-biome map (Fig.1) is derived from the biomes in Olson et al (2001), aggregated to form larger regions.As a result of this setup, each column of **X** includes a single environmental driver for a single biome. Therefore, each environmental driver is represented by a total of seven columns in **X**. Within each column, all elements are zeros except for elements that correspond to a single biome~~ Also note that we standardize (i.e., normalize) each of the environmental driver datasets within each biome and each year before running the regression, as has been done in several previous GIM studies (e.g., Gourdji et al., 2012; Fang and Michalak, 2015). This step means that all of the estimated regression coefficients ($\boldsymbol{\beta}$) have the same units, are independent of the original units on the environmental driver data, and can be directly compared to one another.

~~We also include several constant columns of ones in **X**. These columns are analogous to the intercept in a linear regression. Existing GIM studies always include one or more constant columns within **X** (e. g., Gourdji et al. 2008; Gourdji et al., 2012; Miller et al., 2018). In this study, we specifically use a total of seven constant columns, one for each biome. We also include a constant column for the ocean.~~

2.3 Model selection

~~We further consider non-biospheric fluxes in the **X** matrix, including fossil fuel emissions from the Open-source Data Inventory for Anthropogenic CO_2 monthly fossil fuel emissions (ODIAC2016, Oda et al., 2018) , climatological ocean fluxes from Takahashi et al. (2016), and biomass burning fluxes from the Global Fire Emissions Database (GFED) version 4.1 (Randerson et al., 2018). We only allocate a single column of **X** for fossil fuel, biomass burning, and ocean fluxes, respectively, because these fluxes are not the focus of this study.~~

~~In total, we consider a total of 81 columns for the **X** matrix: 8 constant columns of ones, 70 columns associated with environmental drivers, and 3 columns associated with anthropogenic, ocean, and biomass burning fluxes.~~

~~**2.5 Model selection**~~

We ~~utilize a model selection framework to evaluate which subset of the environmental drivers (i.e., columns of~~ use model selection to decide which environmental driver datasets to include in the analysis of the OCO-2 observations and in the analysis of each TBM using synthetic observations. Model selection ensures that the environmental driver datasets in the regression (**X**) ~~best describe variations in CO_2 fluxes as inferred from the~~ do not overfit the available OCO-2 ~~observations~~data ($\boldsymbol{z}$). The inclusion of additional environmental ~~drivers~~ driver datasets or columns in **X** will al-

ways improve the model-data fit in the regression, but the inclusion of too many ~~variables~~ driver datasets in $\mathbf{X}$ can ~~yield an overfit of the~~ overfit the regression to available OCO-2 ~~observations or can yield unrealistic drift~~ data and result in unrealistic coefficients ($\boldsymbol{\beta}$) ~~(e. g., Zucchini, 2000). Instead of including all environmental drivers in $\mathbf{X}$, we use model selection to decide which set of environmental drivers to include in $\mathbf{X}$~~(e.g., Zucchini, 2000). In addition, model selection indicates which relationships with environmental drivers we can confidently constrain and which we cannot given current OCO-2 observations (e.g., Miller et al., 2018). In this study, we implement a type of model selection known as the Bayesian Information Criterion (BIC~~; Schwarz, 1978) , which has been extensively used in recent GIM studies (e. g., Gourdji et al., 2012; Miller et al. 2013; Fang and Michalak, 2015).~~) (Schwarz et al., 1978) , and various forms of the BIC have been implemented in numerous recent atmospheric inverse modeling studies (e.g., Gourdji et al., 2012; Miller et al., 2013; Fang and Michalak, 2015; Miller et al., 2018; Miller and Michalak, 2020). Using the BIC, we score different combinations of environmental ~~drivers~~ driver datasets that could be included in $\mathbf{X}$ based on how well each combination ~~reproduces the~~ helps reproduce either the real or synthetic OCO-2 observations ~~. We calculate these scores~~ ($\boldsymbol{z}$, Eq. 1). We specifically use an implementation of the BIC from Miller et al. (2018) and Miller and Michalak (2020) that is designed to be computationally efficient for very large satellite datasets. The BIC scores in this implementation are calculated using the following equation~~for the implementation here (Miller et al. 2018; Miller and Michalak, 2020): ~~:

$$BIC = L + p \ln n^* \tag{2}$$

where L is the log likelihood of a particular combination of environmental ~~drivers~~ driver datasets (i.e., columns of $\mathbf{X}$), p is the number of environmental ~~drivers in this~~ driver datasets in a particular combination, and n^* is the effective number of independent observations. This last variable accounts for the fact that not all atmospheric observations are independent, and the model-data residuals can exhibit spatially and temporally correlated errors (Miller et al., 2018). For all simulations here, we use an estimate of n^* for the v9 OCO-2 observations from Miller and Michalak (2020). The first component of Eq. 2 (L) rewards combinations that are a better fit to the ~~observations~~OCO-2 observations ($\boldsymbol{z}$), whereas the second component ~~in Eq. 3~~ of Eq. 2 ($p \ln n^*$) penalizes models with a greater number of columns to prevent overfitting. The best combination of environmental drivers for $\mathbf{X}$ is the combination that receives the lowest ~~score (the Supplement Sect. S3 and Table S2). We implement the BIC using a heuristic branch and bound algorithm (Yadav et al., 2013) to reduce computing time. Miller et al (2018) describes this model selection procedure~~ BIC score. Miller et al. (2018) describes this implementation of the BIC in greater detail, including the specific setup and equations~~for the BIC.~~ .

~~2.6 Posterior uncertainties~~

~~In a GIM, the direct solution to calculate the posterior covariance matrix $\mathbf{V}_s$ (dimensions $m \times m$) can be computed as (~~Note that we run model selection for the OCO-2 data and re-run model selection for each set of synthetic OCO-2 datasets generated using each TBM. As a result, we sometimes select different environmental driver datasets for the analysis using different TBMs. This setup parallels that of Huntzinger et al. (2011) and Fang and Michalak (2015). Furthermore, we use the same set

of environmental driver datasets in each year of the study period (e.g., ~~Saibaba and Kitanidis, 2014; Miller et al. , 2019):~~

$$\mathbf{V}_s = \mathbf{V}_1 + \mathbf{V}_2\mathbf{V}_3\mathbf{V}_2^T$$

$$\mathbf{V}_1 = (\mathbf{Q}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$$

$$\mathbf{V}_2 = \mathbf{V}_1\mathbf{Q}^{-1}\mathbf{X}$$

$$\mathbf{V}_3 = (\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X} - (\mathbf{Q}^{-1}\mathbf{X})^T\mathbf{V_1}\mathbf{Q}^{-1}\mathbf{X})^{-1}$$

2015–2018), a setup that parallels existing GIM studies that use multiple years of atmospheric observations (e.g., Shiga et al., 2018). We estimate different regression coefficients ($\boldsymbol{\beta}$) for each year of the study period, but the actual environmental driver datasets included in the regression does not change from one year to the next. An environmental driver dataset is either selected to be included for all years in a specific biome (based on the BIC scores), or it is not used in any year of the analysis.

2.4 Statistical model for estimating the coefficients ($\boldsymbol{\beta}$)

Once we have chosen a set of environmental driver datasets using model selection, we estimate the coefficients ($\boldsymbol{\beta}$) that relate the real or synthetic OCO-2 observations to these environmental datasets (e.g., Gourdji et al., 2012; Fang and Michalak, 2015) :

$$\hat{\boldsymbol{\beta}} = (h(\mathbf{X})^T\boldsymbol{\Psi}^{-1}h(\mathbf{X}))^{-1}h(\mathbf{X})^T\boldsymbol{\Psi}^{-1}\boldsymbol{z} \quad (3)$$

where ~~the posterior error covariance matrix $\mathbf{V}_s$ is the sum of $\mathbf{V}_1$ and $\mathbf{V}_2\mathbf{V}_3\mathbf{V}_2^T$, and $\mathbf{H}$ is a $n \times m$ matrixdescribing the footprint~~ $\boldsymbol{\Psi}$ ($n \times n$) is a covariance matrix that describes model-data residuals (discussed at the end of this section). Furthermore, the uncertainties in these estimated coefficients can also be estimated using a linear equation (e.g., Gourdji et al., 2008; Fang and Michalak, 2015):

$$\mathbf{V}_{\hat{\boldsymbol{\beta}}} = (h(\mathbf{X})^T\boldsymbol{\Psi}^{-1}h(\mathbf{X}))^{-1} \quad (4)$$

where $\mathbf{V}_{\hat{\boldsymbol{\beta}}}$ is a $p \times p$ covariance matrix.

We test out two different formulations for the covariance matrix $\boldsymbol{\Psi}$ to evaluate the sensitivity of the ~~observations ($\boldsymbol{z}$) to the fluxes ($\boldsymbol{s}$). Note that $\mathbf{V}_1$ is the posterior error covariance matrix in a classic Bayesian inverse model~~ results to the assumptions made about the covariance matrix parameters. In one set of simulations, we model $\boldsymbol{\Psi}$ as a diagonal matrix. The diagonal values characterize model-data errors ($\boldsymbol{\epsilon}$), estimated for the version 9 retrievals from the recent OCO-2 model inter-comparison project (e.g., Crowell et al., 2019). The values have an average standard deviation of 0.98 ppm and range from 0.29 ppm to 4.8 ppm. In a second set of simulations, we use a more complex and more complete formulation of $\boldsymbol{\Psi}$: $\boldsymbol{\Psi} = h(h(\mathbf{Q})^T) + \mathbf{R}$ (e.g., Fang and Michalak, 2015), where $\mathbf{R}$ (~~e. g. , Rodgers, 2000; Brasseur and Jacob, 2017) . $\mathbf{V}_2\mathbf{V}_3\mathbf{V}_2^T$ accounts for the additional uncertainty~~ $n \times n$) characterizes the model-data errors (described above), and $\mathbf{Q}$ ($m \times m$) is a covariance matrix that describes $\boldsymbol{\zeta}$ (the patterns in the fluxes ~~due to the unknown drift~~ that cannot be described by the environmental driver

datasets). This formulation is more complete because it fully accounts for the residuals between z and $\mathbf{X}\boldsymbol{\beta}$. However, it is extremely computationally intensive to estimate the coefficients ($\boldsymbol{\beta}$) ~~.~~ using this complex formulation of $\boldsymbol{\Psi}$. We cannot explicitly formulate this more complex version of $\boldsymbol{\Psi}$ due to its large size and the number of atmospheric model simulations ($h()$) that would be required. As a result, we find the solution to Eq. 3 using this complex version of $\boldsymbol{\Psi}$ by iteratively minimizing the cost function for a geostatistical inverse model (GIM) (Sects. 2.5 and S1), a process that takes approximately two weeks for each year of model simulations in the setup used here.

~~The calculations in Eq. 5 are not computationally feasible for most inverse problems with very large datasets; the matrix sum in $\mathbf{V}_1$ is often too large to invert, and we do not explicitly construct an H matrix or its transpose $\mathbf{H}^T$. Instead, we employ a low-rank approximation of $\mathbf{V}_1$ that circumvents these problems (e. g., Bousserez and Henze, 2018; Wells et al. , 2018). Specifically, we approximate the matrices in $\mathbf{V}_1$ as a low rank update to $\mathbf{Q}$ using a limited number of eigenpairs (i. e., eigenvectors and eigenvalues). Miller et al (2019) and the Supplement Sect.S4 describe the uncertainty quantification in greater detail.~~We use both the simple and complex formulations of $\boldsymbol{\Psi}$ when analyzing the real OCO-2 observations. Both the simple and complex formulations of $\boldsymbol{\Psi}$ yield similar estimates for the coefficients $\boldsymbol{\beta}$, as discussed in the Results & Discussion (Sect. 3.2). When analyzing the 15 TRENDY models, we only use the simple, diagonal formulation of $\boldsymbol{\Psi}$ – because of the prohibitive computational costs that would be required to run the more complex approach for all 15 TRENDY models.

~~**3. Results and Discussion**~~ Note that we estimate the values of $\mathbf{Q}$, the covariance matrix that describes $\boldsymbol{\zeta}$, using an approach known as restricted maximum likelihood estimation (e.g., Mueller et al., 2008; Gourdji et al., 2008, 2010, 2012). In the SI, we discuss the structure of $\mathbf{Q}$ in detail, describe RML, and compare the estimated parameters for $\mathbf{Q}$ against existing studies.

~~**3.1 Connections between CO_2 fluxes and environmental drivers**~~

2.5 Statistical model for estimating CO_2 fluxes using OCO-2 observations

~~A small number of environmental drivers can describe most spatiotemporal variability in~~ To complement the analysis described above, we take an additional step for the real OCO-2 observations of estimating $\boldsymbol{\zeta}$, patterns in the fluxes that cannot be described by the environmental driver datasets, also known as the stochastic component of the fluxes (Eq. 1). This step thereby creates a complete estimate of CO_2 fluxes using OCO-2 observations. This additional step accomplishes two goals. First, the fluxes in $\boldsymbol{\zeta}$ can reveal flux anomalies or patterns that are too complex to quantify using a linear combination of environmental variables and/or can indicate the strengths and shortfalls of the regression. Second, by estimating all components of the CO_2 fluxes ($\mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$), we can better evaluate our inferences using OCO-2 against independent, ground-based observations of CO_2. This independent evaluation is important because OCO-2 observations and the atmospheric transport model (i.e., GEOS-Chem) can contain errors.

We generate a complete estimate of the CO_2 fluxes ~~as estimated in the GIM. In this study, we define spatiotemporal variability as any spatial or temporal patterns~~ ($\mathbf{X}\boldsymbol{\beta}$ + $\boldsymbol{\zeta}$) by minimizing the cost function for a GIM (e.g., Kitanidis, 1986; Michalak et al., 2004; Miller et al., 2020). We describe this process in detail in Supplemental Sect. S1. This process requires two covariance matrices ($\mathbf{R}$ and $\mathbf{Q}$), and we use the same parameters for these covariance matrices as described above in Sect. 2.4. Note that for the setup here, we estimate $\boldsymbol{\zeta}$ at a spatial resolution of 4° latitude by 5° longitude

to match that of GEOS-Chem, and we estimate ζ at a daily temporal resolution to better account for sub-monthly variability in CO_2 fluxes~~that manifest at the daily, (latitude) × (longitude) resolutions of the GEOS-Chem model during the one-year study period (year 2016) . The deterministic model accounts for 89.6% of the variance in the estimated fluxes (Fig. 2a), and the stochastic component conversely accounts for only 10.4% of the flux variance (Fig. 2b).~~. Also note that minimizing the GIM cost function yields the same estimate for the coefficients (β) as in Eq. 3, provided that the covariance matrices in the GIM cost function and in Eq. 3 are identical. The Supplemental Sect. S1 and Miller et al. (2020) describe the process of minimizing the GIM cost function in greater detail.

~~A combination of PAR, daily temperature, and daily precipitation best describe patterns in~~

2.6 Analysis using real observations from OCO-2

For analysis using OCO-2 observations, we employ 10-second averages of the version 9 OCO-2 observations (e.g., Crowell et al., 2019) and include both land nadir- and land glint-mode retrievals. Recent retrieval updates have greatly reduced biases that previously existed between land nadir and land glint observations (O'Dell et al., 2018). Moreover, Miller and Michalak (2020) evaluated the impact of these updated OCO-2 retrievals on the terrestrial CO_2 ~~fluxes in most biomes across the globe(Table 1). PAR is an adept predictor of fluxes across mid-to-high latitudes, whereas a combined set of daily air temperature and daily precipitation are better predictors across tropical biomes~~flux constraint in different regions of the globe; the authors found that the inclusion of both land nadir and land glint retrievals yielded a stronger constraint on CO_2 fluxes relative to using only a single observation type.

~~The deterministic model also includes fossil fuel emissionsfrom ODIAC2016 but not biomass burning fluxes from GFED or ocean fluxesfrom Takahashi et al., (2016). Fossil fuel emissions from ODIAC2016, when passed through the GEOS-Chem model, help describe enough variability in the~~ We also include a column of $\mathbf{X}$ in all simulations using real OCO-2 ~~observations to be selected using the BIC. By contrast, neither~~ observations to account for anthropogenic emissions, ocean fluxes, and biomass burning. This column includes anthropogenic emissions from the Open-Data Inventory for Anthropogenic Carbon Dioxide (ODIAC) (Oda et al., 2018), ocean fluxes from NASA Estimating the Circulation and Climate of the Ocean (ECCO) Darwin (Carroll et al., 2020), and biomass burning fluxes from ~~GFED nor ocean fluxes from Takahashi et al. (2016) help reproduce the~~ the Global Fire Emissions Database (GFED) (Randerson et al., 2018). We estimate a single coefficient or scaling factor (β) for this column. These fluxes are input into the regression at a 4° latitude by 5° longitude spatial resolution to match that of GEOS-Chem. The supplemental Sect. S2 contains greater discussion of these CO_2 sources.

2.7 Analysis using the TBMs

We compare the estimated coefficients (β) from real OCO-2 observations ~~more than the penalty term in the BIC, and these fluxes are therefore not selected using the BIC. Specifically, biomass burning and ocean fluxes may not have been selected for several reasons: either those fluxes are small relative to fossil fuel emissions and biospheric fluxes, the land~~ against simulations using synthetic OCO-2 ~~observations from 2016 are not sensitive to biomass burning and ocean fluxes,~~ observations generated from 15 different TBMs in TRENDY (v8). We list out all of the individual models in the TRENDY comparison in Table S1.

Model outputs from the TRENDY project were provided at a monthly time resolution, and ~~/or the flux patterns in GFED and Takahashi et al. (2016) are not consistent with~~ the spatial resolution varies from one model to another (though many models have a native spatial resolution of either 0.5° latitude-longitude or 1° latitude-longitude). We specifically use TRENDY model outputs from scenario 3 simulations, in which all TBMs are forced with time-varying CO_2, climate, and land use.

We generate synthetic OCO-2 observations using each of these TBM flux estimates. To do so, we first regrid each of the TRENDY model estimates to a spatial resolution of 4° latitude by 5° longitude, the spatial resolution of the GEOS-Chem model. We then run the TRENDY model fluxes through the GEOS-Chem model for years 2015 – 2018 and interpolate the model outputs to the times and locations of the OCO-2 observations. ~~Instead, biomass burning and ocean fluxes are included within the stochastic component of the flux estimate . The Supplement Sect. S6 describes a sensitivity analysis using the BIC that provides further explanation why the deterministic model does not include GFED or ocean fluxes from Takahashi et al. (2016).~~

~~Overall, we only select a limited~~

3 Results & discussion

3.1 Results of model selection

The model selection results highlight the strengths and limitations of using current OCO-2 observations to estimate relationships with environmental driver datasets. We use model selection based on the BIC to determine a set of environmental driver datasets to include in the analysis using OCO-2 observations and using the TBMs. We only select 10 environmental driver datasets when we run model selection on the OCO-2 observations – both when we use environmental driver datasets from the CRUJRA and MERRA-2 products. We are generally able to identify at least one or two key environmental relationships in each biome using total column CO_2 observations (shown on the x-axis of Fig. 3). With that said, we are only able to quantify relationships with these few, salient environmental variables. More detailed environmental relationships within each biome are difficult to discern.

Note that we select a similar number of environmental ~~drivers (12 out of 70, 18%) using model selection. Specifically, we never select more than 3 environmental drivers in any individual biome (Table 1)~~ driver datasets when using synthetic OCO-2 observations that are generated from each of the TBMs. We select anywhere between 8 and 13 environmental driver datasets (an average of 10 datasets) in the analysis using each of the TBMs. This result indicates ~~two likely conclusions. First, a few simple linear relationships may adeptly describe flux variability at the scale and resolution of a global gridded atmospheric model, although the underlying leaf- and organism-level processes are admittedly more complex. Indeed, previous top-down studies (e.g. , Gourdji et al. , 2008, 2012; Fang and Michalak, 2015; Shiga et al. , 2018) also found that simple linear relationships can effectively describe broad spatial and temporal patterns in CO~~consistency between the analysis using real OCO-2 observations and the analysis using synthetic OCO-2 observations generated using CO_2 ~~flux variability across North America and across the globe. Such simple linear relationships allow for a straightforward assessment of the explanatory power of environmental~~

drivers, and make it possible to compare these relationships inferred from atmospheric observations against the relationships used in TBMs (e. g. , Huntzinger et al., 2013; Fang and Michalak, 2015). fluxes from each of the 15 different TBMs.

Second, additional environmental drivers, when run through an atmospheric transport model Overall, we have difficulty detecting the unique contributions of many environmental driver datasets to variability in the OCO-2 observations. This issue is highlighted by an examination of colinearity in the regression model. In a regression, we cannot estimate different coefficients (β) for two predictor variables (i.e., columns of $\mathbf{X}$) that are identical or nearly identical; the regression cannot be used to estimate unique coefficients because the predictor variables themselves are not unique. In regression modeling, this phenomenon is known as colinearity. The coefficients (β) estimated for colinear variables are often unrealistic, and interpolated to the times and the standard errors or uncertainties in those coefficients ($\mathbf{V}_{\hat{\beta}}$) are often unexpectedly large (e.g., Ramsey and Schafer, 2012). Model selection is one way to reduce or remove colinearity; colinear variables, by definition, do not contribute unique information to a regression and are therefore rarely selected using a model selection approach like the BIC. One common method for detecting colinearity is to estimate the correlation coefficient (r) between different columns of $\mathbf{X}$; a value greater than ~0.55 can indicate the presence of colinearity (e.g., Ratner, 2012).

We find substantial colinearity in the regression analysis (Fig. 2). This colinearity likely plays an important role in the model selection results, in addition to errors in the OCO-2 observations and errors in the GEOS-Chem model; it represents and important but potentially overlooked challenge in relating satellite-based CO_2 observations to patterns in environmental drivers. The environmental driver datasets are passed through the GEOS-Chem model ($h()$) and interpolated the locations of OCO-2 observations , are not sufficiently unique to parse out their differing relationships with as part of the regression ($h(\mathbf{X})$, Eqs. 1 and 3). In other words, these driver datasets are input into GEOS-Chem in place of a traditional CO_2 fluxes. Model selection ensures that we only include environmental drivers that contribute unique information to the flux estimate and do not overfit the flux estimate. This step is necessary so that the environmental driver datasets can be directly compared against patterns in the OCO-2 observations. If multiple environmental drivers are highly correlated or colinear, then the inclusion of more than one of these drivers will not contribute unique information. As a result, we are unable to quantify a larger number of environmental driver relationships using OCO-2The driver datasets (i.e., columns of $\mathbf{X}$) that we use in the regression are generally unique from one another (i.e., have unique spatial and temporal patterns). However, the differences among many driver datasets disappear once those datasets have been passed through GEOS-Chem. Fig. 3 illustrates an example of air temperature and PAR. In most of the biomes, there is aweak correlation ($R < 0.4$; left column) between 2-m air temperature and PAR; however, the correlation is much stronger ($R > 0.8$; right column) when these environmental drivers are passed through an atmospheric model ($h(\mathbf{X})$). A larger number of environmental drivers is not selected due to this high level of correlation or collinearity among the columns in $h(\mathbf{X})$. This collinearity, not errors in the 2 displays the correlation coefficients (r) among environmental driver datasets from MERRA-2 for temperate forests, both before (Fig. 2a) and after (Fig. 2b) those driver datasets have been passed through GEOS-Chem and interpolated to the OCO-2 retrievals or atmospheric model, appears to be a limiting factor in the model selection results.

3.1.1 PAR shows stronger explanatory power than temperature or precipitation in mid-to-high latitudes

PAR is selected for four biomes: temperature forests, temperate grasslands, boreal forests and tropical forests (Table 1). In the middle and high latitudes, PAR, rather than temperature or precipitation, appears to be a better proxy for seasonal patterns in observations. The correlation coefficients increase substantially from the former to the latter case. This colinearity is independent of errors in the OCO-2 observations and indicates a hard limit on the number of relationships with environmental driver datasets (i.e., coefficients) that we can quantify in the regression. In other words, model selection results are at least partially limited by the limited sensitivity of OCO-2 observations to variations in these environmental driver datasets – either due to atmospheric smoothing and/or due to limitations in the availability of OCO-2 observations in some regions of the globe. This limitation is in addition to the uncertainties due to errors in the OCO-2 observations and the GEOS-Chem model, which also have a critical impact on inferences about CO_2 fluxes (Figs. 4a, b and S3a-f). using OCO-2 observations (e.g., Chevallier et al., 2007, 2014; Miller et al., 2018).

3.2 Environmental relationships inferred using observations from OCO-2

We are able to quantify the relationships between OCO-2 observations and several key environmental driver datasets. Figs. 3 and 4 displays the estimated coefficients from the regression analysis using observations from OCO-2. Across extratropical biomes, PAR is the most commonly selected variable. This result reflects the fact that light availability is likely an important a key factor that drives CO_2 flux variability in mid-to-high latitudes (e. g., Fang and Michalak, 2015; Baldocchi et al., 2017). The β values for PAR indicate a strong to moderate negativecorrelation with estimated CO_2 fluxes, suggesting (e.g., Fang and Michalak, 2015; Baldocchi et al., 2017). As expected, the estimated coefficients for PAR are negative, indicating that an increase (or decrease) in PAR in the model is associated with a decrease (or increase) in NEE, and indicating an increase (or decrease) in carbon uptake; this β value is larger in boreal and temperate forests relative to grasslands, indicating a stronger relationship between PAR and net biosphere . Note that in this study, negative values for NEE refer to CO_2 fluxes in those biomes (Table 1; Figs. 5a and S3d-f).

Indeed, previous studies also indicate that PAR and similar environmental drivers (e.g., shortwave radiation) are closely associated with uptake while positive values refer to net CO_2 fluxes. For example, a top-down study of North America (Fang and Michalak, 2015) found that shortwave radiation is more adept than other environmental variables in reproducing spatiotemporal variability of NEE, particularly across the growing season. Moreover, several site-level studies have reached parallel conclusions (e.g., Mueller et al., 2010; Yadav et al., 2010); these studies indicated that PAR is strongly correlated with photosynthesis, consistent with current mechanistic understandings of the light limitation on photosynthesis (e.g., Gough et al., 2007). release to the atmosphere.

3.1.2 Drought is likely associated with flux variations across tropical forests

A composite of PAR, scaled temperature, and daily precipitation adeptly describe variability in COBy contrast, precipitation and scaled temperature are the most commonly selected environmental driver datasets across tropical biomes. The magnitude of the coefficients for each of these two variables is similar in most biomes, indicating that patterns in both have similarly important associations with patterns in CO_2 fluxes across tropical forests (Figs. 4c and 4d), as seen through the OCO-2 observations. PAR in tropical forests is usually a function of the presence or absence of clouds (e.g., Baldocchi et al., 2017; Zeri

et al., 2014); cloudiness is also associated with rainfall. Therefore, low PAR over tropical forests is likely an indicator of cloud presence and rainfall. A positive β estimated for PAR suggests that a decrease in PAR, indicative of enhanced precipitation, is associated with 380 increased carbon uptake. Furthermore, the negative β value fluxes. Specifically, a negative coefficient assigned to scaled temperature (the Supplement Sect. S2) implies indicates that an increase in air temperature , which often exceeds optimal temperature over tropical forests, is associated with increased carbon uptake when air temperatures are cool and reduced carbon uptake .

Recent studies (e.g., Jiménez-Muñoz et al., 2016; Liu et al., 2017; Palmer et al., 2019) indicated that tropical droughts associated with the 2015-2016 El Niño events likely resulted in above average carbon release. Indeed, the combination of high values of PAR, high air temperature, and low precipitation may be a manifestation of these drought patterns.

Indeed, multiple lines of evidence indicate that drought is associated with diminished carbon uptake in tropical forests when air temperatures are hot; the scaled temperature function has the shape of an upside down parabola, and temperature thus has a different association with CO_2 fluxes depending upon whether the air temperature is above or below the optimal temperature for photosynthesis (e.g., Phillips et al. , 2009; Brienen et al., 2015; Baccini et al., 2017). For example, Gatti et al (2014) suggested that a suppression of photosynthesis during tropical drought may cause a reduction in carbon uptake. Brienen et al (2015) added that tropical drought is often associated with higher-than-normal temperature , which may further contribute to reducing gross primary production (GPP) and carbon uptake . Overall, this GIM study supports the conclusion that environmental conditions indicative of drought are associated with net carbon emissions from tropical forests.

3.1.3 CO_2 fluxes, as inferred from OCO-2, are closely correlated with temperature and precipitation in tropical grasslands

Temperature and precipitation closely correlate with variability in CO_2 fluxes across tropical grasslands (Figs. S3g and S3j). This result suggests that heat and water availability are likely associated with carbon fluxes across this biome.

A negative β value for precipitation indicates Fig. S1). Indeed, high temperatures in the tropics often exceed the optimal temperature for photosynthesis (e.g., Baldocchi et al., 2017), which reduces carbon uptake (e.g., Doughty and Goulden, 2008) . Furthermore, negative coefficients for precipitation indicate that an increase in precipitation is associated with an increase in carbon uptake, which is in line with current knowledge that water availability facilitates photosynthesis across seasonal to annual temporal scales, especially in arid or semi-arid regions . In addition, a negative β value for scaled temperature (the Supplement Sect. S2) indicates that an increase in air temperature is associated with a reduction in carbon uptake. Specifically, high temperatures in the tropics often exceed the optimal temperature for photosynthesis (e. g. , Baldocchi et al., 2017), which can suppress GPP (e.g., Doughty and Golden, 2008). Overall, a combined set of air temperature and precipitation adeptly describes COsemiarid regions (e.g., Gatti et al., 2014; Jung et al., 2017).

In addition to this regression analysis, we use a GIM to estimate the stochastic component of the fluxes (ζ, Eq. 1) – patterns in the fluxes that are implied by the OCO-2 observations but do not match any existing environmental driver dataset. To this end, Fig. 5 shows the mean contribution of each environmental driver variable and the stochastic component to the GIM across years 2015 – 2018 using MERRA-2 for the environmental driver datasets. The magnitude of the stochastic component in this plot is small relative to the contribution of different environmental variables and relative to the contribution of anthropogenic

sources. Furthermore, the stochastic component contains very diffuse spatial patterns, and these very broad patterns do not imply any clear deficiency in the other components of the GIM. For example, the regression component of the GIM ($\mathbf{X}\hat{\beta}$) accounts for 89.6% of the variance in the estimated fluxes, and the stochastic component conversely accounts for only 10.4% of the flux variance. Furthermore, the regression component, when passed through the GEOS-Chem model, matches OCO-2 observations nearly as well as the full posterior flux estimate (Figs. S2 and S13). This result shows that a limited number of environmental driver datasets can adeptly reproduce broad patterns in CO_2 ~~flux variability in tropical grasslands, rendering it a net source in year 2016.~~ fluxes across continental and global spatial scales but reinforces the conclusion that current OCO-2 observations are not sufficient to disentangle more complex environmental relationships.

~~3.2 Estimated biospheric flux totals for different global regions~~

~~We estimate a global terrestrial biospheric CO_2 budget of -1.73 $\pm$ 0.53 GtC (Uncertainties listed are the 95% confidence interval. The Supplement Sect. S5 provides detail on the posterior uncertainty estimate for biospheric fluxes.). Among the seven biomes, middle to high latitudes (primarily temperate, boreal and tundra biomes) act as a significant carbon sink; tropical biomes are a net source; desert and shrubland regions play a small, neutral role (Table 2). Note that we subtract flux patterns that map onto fossil fuels ($\mathbf{X}\beta$, Fig. 5d) from the posterior flux estimate (s, Fig. 2c) to obtain an estimate for biospheric fluxes (including terrestrial NEE and biomass burningfluxes). We estimate a β value of 1.09 $\pm$ 0.05 (95% confidence interval) for the fossil fuel emissions from ODIAC2016~~In all of the simulations using OCO-2 observations, we estimate a scaling factor (β) for anthropogenic, biomass burning, and ocean fluxes of near one, indicating that ~~the overall global magnitude of ODIAC2016 is consistent with OCO-2 observations. We therefore assume that ODIAC2016 is a reasonable global estimate for fossil fuel emissions.~~ these source types have a magnitude that is broadly consistent with atmospheric observations. Specifically, the estimated scaling factor estimated ranges from 0.97 to 1.05, depending upon the year and simulation. Note, however, that we estimate a single scaling factor for all of these source types combined and are unable to confidently constrain separate scaling factors for each source, a topic discussed in greater detail in the Supplemental Sect. S2.

~~These flux totals are~~ Note that the inferences described here are also broadly consistent with ~~a recent MIP of different inverse models that assimilate OCO-2 observations(Crowell et al. , 2019). The inverse modeling teams that participated in the MIP employed different transport models, inverse modeling approaches, and prior flux assumptions. The total global terrestrial biospheric flux, averaged across all models, was -1.4 $\pm$ 0.7 GtC for the year of 2016. The MIP fluxes assimilate v7 of land nadir-mode XCO2 retrievals; unlike this study in which we use v9 of land nadir- and glint-mode retrievals. In spite of this difference, the averaged global flux from the MIP study and the estimate reported here are very similar.~~

~~In order to provide an additional comparison with the MIP results, we group the estimated fluxes into TRANSCOM land regions (Gurney et al., 2002)~~independent, ground-based atmospheric observations. ~~We split the classic TRANSCOM regions at the Equator to avoid regions that encompass parts of both the northern and southern hemisphere, as in Crowell et al (2019). In most of the regions,~~ We specifically model atmospheric CO_2 using fluxes estimated from the GIM and compare against regular aircraft observations, campaign data from the Atmospheric Tomography Mission (ATom; Wofsy et al., 2018), and observations from Total Carbon Column Observing Network (TCCON; Wunch et al., 2011). In most instances, the model result matches the observations to within the errors specified in the inverse model (i.e., to within the errors specified in the $\mathbf{R}$ covariance matrix),

and the model-data comparisons do not exhibit any obvious seasonal biases. Furthermore, we also model XCO_2 using the outputs of the regression analysis ($\mathbf{X}\hat{\beta}$), and these outputs also show good agreement with OCO-2 observations (Fig. S13). The Supplemental Sect. S4, Figs. S2-S13, and Tables S2-S3 describe these comparisons in greater detail.

3.3 Comparison between inferences from OCO-2 and TBMs

The environmental relationships (i.e., coefficients) estimated for the TBMs show a substantial range (Figs. 3 and 4); this spread highlights uncertainties in state-of-the-art TBMs and indicates that there is an opportunity to help inform these relationships using atmospheric CO_2 observations. On one hand, we are only able to infer a limited number of environmental relationships using current observations from OCO-2, and this fact limits the extent to which we can inform TBM development using available space-based CO_2 observations. On the other hand, we can infer relationships with several key environmental drivers (e.g., Fig. 3), and TBMs disagree on relationships with even these key drivers. This result thus indicates both the limitations of this analysis but also its strengths. Specifically, Figs. 3 and 4 graphically displays these results from the regression analysis – the coefficients estimated using OCO-2 observations compared to those estimated from the TRENDY models. The coefficients from the OCO-2 analysis are almost always within the range of those estimated using the ensemble of TBMs. With that said, the coefficients estimated for many of the ~~fluxes estimated using the GIM are very~~ TBMs are far from the value estimated using OCO-2, implying that observations from OCO-2 can be used to inform the relationships within numerous individual models. Note that in Figs. 3 and 4, the x-axis is ordered based upon the environmental driver variables that are selected using OCO-2, and we show the estimated coefficients for TBMs in which the listed environmental driver variable is also chosen using model selection. Furthermore, the coefficients shown for the TBMs in Figs. 3 and 4 are calculated using environmental driver datasets from CRUJRA. Fig. S14 displays the results for the TBMs using environmental driver data from MERRA-2, and the results look similar to those ~~reported in the MIP~~ using CRUJRA.

We specifically find large differences between the analysis using OCO-2 and the TBMs for relationships with precipitation. The relationships between precipitation are arguably more uncertain within the TBMs than the relationships with other environmental variables (Fig. ~~6) ; however, the fluxes estimated here are significantly different in a limited number of regions (e. g. , tropical Australia and northern tropical Africa), a possible reflection of differences between the v9 and v7~~ 3a) and are more uncertain in tropical biomes than temperate ones. This statement is particularly apparent when we examine the coefficient of variation for each relationship (Fig. 3b). The coefficient of variation is a measure of the uncertainty relative to the magnitude of the mean, and Fig. 3b shows the standard deviation in the coefficients from the 15 TBMs divided by the mean coefficient from the ensemble of TBMs. In addition, the TBMs are evenly split on whether the relationship with precipitation is positive or negative across tropical biomes, and our analysis using OCO-2 ~~retrievals (O'Dell et al., 2018; Miller et al., 2019). For example, we estimate a smaller CO~~observations agrees with models that estimate a negative relationship (i.e., precipitation is associated with greater CO_2 ~~source for northern tropical Africa relative to the MIP study. However, previous studies (e.g., Wang et al., 2019) indicated that existing satellite-based estimates of~~ uptake). There is substantial disagreement on the magnitude of this relationship, even among models that yield a negative relationship; the estimate using OCO-2 observations falls in the mid-range of these TBMs for both tropical biomes.

More broadly, the TBMs simulate very different water cycling through each ecosystem, in spite of the fact that each model uses the same precipitation inputs from CRUJRA. These broader differences in water cycling within the TBMs may help explain the large uncertainties in the relationships between CO_2 ~~fluxes for this region may be too high. OCO-2 collects far more observations across northern Africa during the dry season than the wet season due to persistent cloudiness in~~ and precipitation, and highlights an important source of uncertainty within these models. Specifically, we find that estimated evapotranspiration (ET) across the ~~wet season, and~~ TBMs differs by almost a factor of three among models in some seasons and biomes, and annual ET ranges from 375 mm to 700 mm over North Hemispheric tropical grasslands (Fig. 6a), and from 530 mm to 1010 mm over North Hemispheric tropical forests (Fig. 6b). These large differences in ET estimates reinforce the very different responses of tropical ecosystems in these models (both tropical forests and tropical grasslands) to precipitation inputs.

Indeed, existing studies have ~~postulated that this difference in data availability may be to blame for a high bias in~~ indicated large uncertainties in the responses of tropical forests to water availability (e.g., Restrepo-Coupe et al., 2016) and have offered several possible explanations. Soil depths and rooting distribution are particularly challenging to model in tropical ecosystems, yielding uncertainties in the relationship between water availability and CO_2 fluxes ~~estimated from OCO-2 (Crowell et al. 2019; Wang et al. 2019).~~ (e.g., Baker et al., 2008; Poulter et al., 2009). For example, Poulter et al. (2009) argued that current TBMs tend to underestimate soil depths in tropical forests, which are critical to guarantee soil water access and to accurately simulate dry-season photosynthesis in TBMs. The treatment of irrigation and other land management practices also differs among models and creates further uncertainty (e.g., Le Quéré et al., 2018; Pan et al., 2020). To complicate matters, the role of precipitation in carbon dynamics can vary depending on broader environmental conditions and the time scales considered (e.g., Baldocchi et al., 2017). For example, excess precipitation is associated with limited light availability in regions like the humid tropics and can raise the water table to a level that inhibits respiration. With that said, short-term rain events have been shown to boost respiration (e.g., Baldocchi, 2008).

~~The fluxes estimated here are also broadly consistent with aircraft-based *in situ* CO_2 observations, a topic discussed in the Supplement Sect. S7.~~

~~3.3 Estimated posterior uncertainties~~

~~The posterior uncertainties for individual biomes range from 0.25 to 0.76 GtC yr^{-1}. Estimated fluxes for tropical forests have higher uncertainties than any other biome (0.76 GtC yr^{-1}), likely a consequence of poor observational coverage due to persistent cloudiness. By contrast, a large number of good-quality~~ Like precipitation, relationships with PAR are also highly uncertain in the simulations using TBMs. Most models yield relationships with the same sign, but those relationships vary widely in magnitude. By contrast, results using OCO-2 ~~retrievals provides robust constraints over temperate forests, yielding a small posterior uncertainty (0.27 GtC yr^{-1}) in the estimated flux~~observations are very similar to the ensemble mean of the TBMs. This result is particularly interesting given that the individual TBMs do not show consensus with one another. The differences among the TBMs likely stem from the fact that these TBMs exhibit widely varying seasonal cycles and peak growing season uptake across extratropical biomes. For example, in temperate forests (e.g., Fig. S17), the maximum monthly carbon uptake differs by a factor of eight among the TBMs, and a handful of TBMs estimate a very different seasonal cycle than the bulk of the TBMs with maximum uptake during the middle of the growing season.

~~It is important to note that the posterior uncertainties calculated in most classical Bayesian or geostatistical inverse models account for many but not all possible sources of uncertainty. For example, the posterior uncertainties presented here account for the sparsity of~~ In contrast to the discussion of precipitation and PAR, existing TBMs yield much better agreement on the relationships between CO_2 and scaled temperature (Fig. 3). In tropical biomes, nearly all TBMs agree on the sign of the relationship, and the estimates using OCO-2 observations are within the range of those estimated using TBMs. Interestingly, the uncertainty bounds on the coefficients estimate using OCO-2 are not that much smaller than the range of coefficients from the ensemble of TBMs, both for tropical grasslands and especially for tropical forests. This result points to relatively good consensus in modeled relationships with temperature for tropical grasslands and forests – both using TBMs and OCO-2 observations~~, random observational or atmospheric transport errors, and uncertainties due to uncertain drift coefficients (β)~~ ~~(~~. However, it also indicates that atmospheric observations from OCO-2 potentially have less opportunity to inform these relationships than for precipitation or PAR where TBMs do not show consensus.

The comparisons described above are largely from biomes centered in the tropics and mid-latitudes and include few comparisons for high latitude biomes (e.g., ~~Kitanidis and Vomvoris, 1983; Michalak et al., 2004). However, these calculations do not fully account for bias-type errors: regional- or continental-scale biases in the OCO-2 observations, biases in modeled atmospheric convection (e. g. , Basu et al. , 2018; Schuh et al., 2019), or biases in modeled interhemispheric transport, among other possible biases. Most classical Bayesian and geostatistical inverse models assume that the observational or model errors are Gaussian with a mean of zero (e. g. , Kitanidis and Vomvoris. 1983; Michalak et al. , 2004; Tarantola, 2005) , making it challenging to account for the types of biases listed above. As a result, the posterior uncertainties estimated in this study are typically smaller than~~ the boreal forest or tundra biomes). For example, we do not select any environmental driver variables for the ~~range of flux estimates produced from the recent MIP study (Fig. 6; Crowell et al., 2019).~~ tundra biome using OCO-2 and only select PAR in boreal forests. OCO-2 observations are sparse across high latitudes both due to the lack of sunlight in winter and due to frequent cloud cover in many high latitude regions. We also only select PAR in boreal forests in simulations using two of the 15 TBMs. This result also reflects the limited availability of OCO-2 observations over high latitude regions; for the analysis here, we create synthetic OCO-2 observations using each TBM and apply model selection to each of these synthetic OCO-2 datasets. Hence, the sparsity of OCO-2 observations not only affects the model selection results using real OCO-2 observations but also affects the analysis shown in Figs. 3 and 4 using the TBMs. The fact that PAR is selected for so few TBMs is not a reflection on the important role of PAR across the boreal forest in many TBMs.

Note that the analysis described above is based upon the mean relationships that we infer for years 2015–2018. We also explored how these relationships in the models vary during El Niño (2015–2016) and non-El Niño years (2017–2018) (Fig. 4). The relationships that we estimate do not fundamentally change between El Niño and non-El Niño years and neither does the spread among the models. This result indicates two conclusions: (1) there is not a fundamental shift in these relationships between El Niño versus non-El Niño years, suggesting that it is not the change in environmental relationships but the change in environmental variables themselves that correlate with the change in flux estimates; and (2) the uncertainties in the relationships, as estimated by the TBMs, are not higher in El Niño versus non-El Niño years. With that said, the magnitude of the estimated coefficient does change in some models between El Niño and non-El Niño years; the changes in the coefficients

are generally less than 50% in most models, and the models do not show a consistent direction of change between El Niño and non-El Niño years.

~~4. Conclusions~~

4 Conclusions

In this study, we ~~adapt the geostatistical approach to inverse modeling for global satellite observations of~~ use four years of observations from OCO-2 and a top-down statistical framework to evaluate the relationships between patterns in atmospheric CO_2 ~~, and evaluate the extent to which we can use these observations to make connections between~~ observations and patterns in environmental driver datasets that are commonly used in modeling the global carbon cycle. We are able to quantify a limited number of these environmental relationships using observations from OCO-2. In spite of these limitations, we are still able to identify relationships with a small number of salient environmental drivers datasets, and state-of-the-art TBMs do not show consensus on some of these key relationships, indicating an opportunity to inform these relationships using atmospheric CO_2 ~~fluxes and environmental drivers. We find that A simple combination of environmental drivers can adeptly describe patterns in CO_2 fluxes across different biomes of the globe, as seen through observations from the~~ observations.

We subsequently compare inferences using OCO-2 ~~satellite; PAR is an adept predictor of fluxes across mid-to-high latitudes, whereas a combination of daily air temperature and daily precipitation shows strong explanatory power across tropical biomes; A larger number of environmental drivers is not selected because many drivers are correlated or colinear when passed through an atmospheric model and averaged across a total atmospheric column. This high collinearity, not errors in the~~ against inferences from 15 state-of-the-art TBMs that have model outputs available for the same set of years. For the broad regions and timespan explored in this study, we find negative relationships between patterns in OCO-2 ~~retrievals or atmospheric model, appears to be a limiting factor in using satellite observations to connect~~ observations and patterns in precipitation; this result agrees with half of the TBMs, which do not show consensus on relationships with precipitation. By contrast, TBMs exhibit much greater skill in describing relationships with scaled temperature, as implied by the relatively good agreement among TBMs. In fact, the uncertainties in the temperature relationship across tropical biomes, as estimated using OCO-2 observations, is nearly as large as the range of estimates using TBMs.

More broadly, state-of-the-art TBMs disagree on the contribution of individual biomes to the global carbon balance, a result highlighted in several studies (e.g., Poulter et al., 2014; Sitch et al., 2015; Ahlström et al., 2015; Piao et al., 2020). In order to reduce these uncertainties, scientists will likely need to reconcile differences in the environmental processes that drive these CO_2 ~~fluxes with environmental drivers; We estimate a global terrestrial biospheric budget of -1.73 ± 0.53 GtC in year 2016, in close agreement with recent inverse modeling studies that use~~ flux estimates. Existing studies have used in situ atmospheric observations to help quantify and evaluate these relationships across the extratropics (e.g., Fang and Michalak, 2015; Hu et al., 2019). However, this task is much more challenging across regions of the globe with sparse in situ observations, including most of the tropics. In spite of the limitations described in this study, the advent of

satellite-based CO_2 observations like those from OCO-2 ~~retrievals as observational constraints .~~ provide a new opportunity to constrain these environmental relationships and thereby provide unique atmospheric constraints on the global carbon cycle.

Data availability. The version 9 of 10-s average OCO-2 retrievals are available at ftp://ftp.cira.colostate.edu/ftp/BAKER/; data information of the OCO-2 MIP is available at https://www.esrl.noaa.gov/gmd/ccgg/OCO2/; data information of the ObsPack data product is available at http://www.esrl.noaa.gov/gmd/ccgg/obspack/.

Author contributions. Z.C. and S.M.M. designed the study, analyzed the data, and wrote the manuscript. S.S., P.F., V.B., D.S.G., V.H., A.J., E.J., E.K., S.L., D.L.L., P.C.M., J.R.M., J.E.M.S.N., B.P., H.T., A.J.W, and S.Z. provided TRENDY model flux estimates. All authors reviewed and edited the paper.

Competing interests. The authors declare they have no competing interests.

Acknowledgements. Financial support for this research has been provided by NASA ROSES grant no. 80NSSC18K0976. We thank ~~David Baker and Andrew Jacobson for their~~ Kim Mueller and Anna Michalak for their feedback on the research; David Baker for his help with the ~~OCO-2 retrievals and the NASA MIP products. We thank~~ XCO_2 data; Colm Sweeny and Kathryn McKain for their help with aircraft datasets from the NOAA/ESRL Global Greenhouse Gas Reference Network~~. We also thank~~ ; John Miller, Luciana Gatti, Wouter Peters and Manuel Gloor for their help with the aircraft data from the INPE ObsPack data product~~. All modeling and analysis was~~ ; Steven Wofsy, Kathryn McKain, Colm Sweeny, and Róisín Commane for their help with ATom aircraft datasets; and Debra Wunch for her help with the TCCON datasets. Daven Henze's work is supported by NOAA grant no. NA16OAR4310113. Daniel Goll's work is supported by the ANR CLAND Convergence Institute. The data analysis and inverse modeling were performed on the NASA Pleiades Supercomputer.

References

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., et al.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink, Science, 348, 895–899, https://doi.org/10.1126/science.aaa1668, 2015.

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., and Da Rocha, H. R.: Seasonal drought stress in the amazon: Reconciling models and observations, J. Geophys. Res. Biogeosciences, 114, 1–10, https://doi.org/10.1029/2007JG000644, 2008.

Baldocchi, D.: 'Breathing' of the Terrestrial Biosphere: Lessons Learned from a Global Network of Carbon Dioxide Flux Measurement Systems, Aust. J. Bot., pp. 1–82, 2008.

Baldocchi, D., Chu, H., and Reichstein, M.: Inter-annual variability of net and gross ecosystem carbon fluxes: A review, Agric. For. Meteorol., 249, 520–533, https://doi.org/10.1016/j.agrformet.2017.05.015, 2017.

Bastos, A., Friedlingstein, P., Sitch, S., Chen, C., Mialon, A., Wigneron, J.-P., Arora, V. K., Briggs, P. R., Canadell, J. G., Ciais, P., et al.: Impact of the 2015/2016 El Niño on the terrestrial carbon cycle constrained by bottom-up and top-down approaches, Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 20170 304, https://doi.org/10.1098/rstb.2017.0304, 2018.

Byrne, B., Liu, J., Lee, M., Baker, I., Bowman, K. W., Deutscher, N. M., Feist, D. G., Griffith, D. W., Iraci, L. T., Kiel, M., Kimball, J. S., Miller, C. E., Morino, I., Parazoo, N. C., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Velazco, V. A., Wennberg, P. O., and Wunch, D.: Improved Constraints on Northern Extratropical CO_2 Fluxes Obtained by Combining Surface-Based and Space-Based Atmospheric CO_2 Measurements, J. Geophys. Res. Atmos., 125, https://doi.org/10.1029/2019JD032029, 2020a.

Byrne, B., Liu, J., Bloom, A. A., Bowman, K. W., Butterfield, Z., Joiner, J., Keenan, T. F., Keppel-Aleks, G., Parazoo, N. C., and Yin, Y.: Contrasting regional carbon cycle responses to seasonal climate anomalies across the east-west divide of temperate North America, Global Biogeochem. Cycles, pp. 1–26, https://doi.org/10.1029/2020gb006598, 2020b.

Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Landschützer, P., Lauderdale, J. M., Liu, J., Manizza, M., Naviaux, J. D., Rödenbeck, C., Schimel, D. S., Van der Stocken, T., and Zhang, H.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean p CO_2 and Air-Sea CO_2 Flux , J. Adv. Model. Earth Syst., 12, 1–28, https://doi.org/10.1029/2019ms001888, 2020.

Chen, N., Zhu, J., Zhang, Y., Liu, Y., Li, J., Zu, J., and Huang, K.: Nonlinear response of ecosystem respiration to multiple levels of temperature increases, Ecol. Evol., 9, 925–937, https://doi.org/10.1002/ece3.4658, 2019.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007375, 2007.

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Philippe, B.: Toward robust and consistent regional CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 31, 1–8, https://doi.org/10.1002/2013GL058772.Received, 2014.

Crisp, D.: Measuring atmospheric carbon dioxide from space: The GOSAT and OCO-2 missions, Opt. InfoBase Conf. Pap., 2, https://doi.org/10.1364/e2.2011.ewc6, 2015.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., Chatterjee, A., Crisp, D., Eldering, A., Jones, D. B., McKain, K., Miller, J., Nassar, R., Oda, T., O'Dell, C., Palmer, P. I., Schimel, D., Stephens, B., and Sweeney, C.: The

2015–2016 Carbon Cycle As Seen from OCO-2 and the Global In Situ Network, Atmos. Chem. Phys. Discuss., pp. 1–79, https://doi.org/10.5194/acp-2019-87, 2019.

Dargaville, R. J., Heimann, M., McGuire, A. D., Prentice, I. C., Kicklighter, D. W., Joos, F., Clein, J. S., Esser, G., Foley, J., Kaplan, J., Meier, R. A., Melillo, J. M., Moore, B., Ramankutty, N., Reichenau, T., Schloss, A., Sitch, S., Tian, H., Williams, L. J., and Wittenberg, U.: Evaluation of terrestrial carbon cycle models with atmospheric CO_2 measurements: Results from transient simulations considering increasing CO_2, climate, and land-use effects, Global Biogeochem. Cycles, 16, https://doi.org/10.1029/2001gb001426, 2002.

Dayalu, A., William Munger, J., Wofsy, S. C., Wang, Y., Nehrkorn, T., Zhao, Y., McElroy, M. B., Nielsen, C. P., and Luus, K.: Assessing biotic contributions to CO_2 fluxes in northern China using the Vegetation, Photosynthesis and Respiration Model (VPRM-China) and observations from 2005 to 2009, Biogeosciences, 15, 6713–6729, https://doi.org/10.5194/bg-15-6713-2018, 2018.

Doughty, C. E. and Goulden, M. L.: Seasonal patterns of tropical forest leaf area index and CO_2 exchange, Journal of Geophysical Research: Biogeosciences, 113, https://doi.org/10.1029/2007JG000590, 2008.

Eldering, A., O'Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., Avis, C., Braverman, A., Castano, R., Chang, A., Chapsky, L., Cheng, C., Connor, B., Dang, L., Doran, G., Fisher, B., Frankenberg, C., Fu, D., Granat, R., Hobbs, J., Lee, R. A., Mandrake, L., McDuffie, J., Miller, C. E., Myers, V., Natraj, V., O'Brien, D., Osterman, G. B., Oyafuso, F., Payne, V. H., Pollock, H. R., Polonsky, I., Roehl, C. M., Rosenberg, R., Schwandner, F., Smyth, M., Tang, V., Taylor, T. E., To, C., Wunch, D., and Yoshimizu, J.: The Orbiting Carbon Observatory-2: First 18 months of science data products, Atmos. Meas. Tech., 10, 549–563, https://doi.org/10.5194/amt-10-549-2017, 2017.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data, Atmos. Meas. Tech., 12, 2341–2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

Fang, Y. and Michalak, A. M.: Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers, Global Biogeochem. Cycles, pp. 1199–1214, https://doi.org/10.1002/2014GB004832, 2015.

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, https://doi.org/10.5194/bg-11-6985-2014, 2014.

Feng, L., Palmer, P. I., Parker, R. J., Deutscher, N. M., Feist, D. G., Kivi, R., Morino, I., and Sussmann, R.: Estimates of European uptake of CO_2 inferred from GOSAT XCO_2 retrievals: sensitivity to measurement bias inside and outside Europe, Atmospheric chemistry and physics, 16, 1289–1302, https://doi.org/10.5194/acp-16-1289-2016, 2016.

Forkel, M., Carvalhais, N., Rödenbeck, C., Keeling, R., Heimann, M., Thonicke, K., Zaehle, S., and Reichstein, M.: Enhanced seasonal CO_2 exchange caused by amplified plant productivity in northern ecosystems, Science (80-.)., 351, 696–699, https://doi.org/10.1126/science.aac4971, 2016.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al.: Climate–Carbon Cycle Feedback Analysis Results from the C4MIP, J. Clim., 2006.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., DBakker, O. C., Canadell1, J. G., Ciais1, P., Jackson, R. B., Anthoni1, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Goldewijk, K. K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E., Nakaoka, S. I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian,

H., Tilbrook, B., Tubiello, F. N., Van Der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global carbon budget 2019, Earth Syst. Sci. Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Gatti, L. V., Gloor, M., Miller, J. B., Doughty, C. E., Malhi, Y., Domingues, L. G., Basso, L. S., Martinewski, A., Correia, C. S., Borges, V. F., Freitas, S., Braz, R., Anderson, L. O., Rocha, H., Grace, J., Phillips, O. L., and Lloyd, J.: Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements, Nature, 506, 76–80, https://doi.org/10.1038/nature12957, 2014.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Clim., 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, J. Geophys. Res., 113, 1–15, https://doi.org/10.1029/2007jd009733, 2008.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E., and Michalak, A. M.: Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: A synthetic data study, Atmos. Chem. Phys., 10, 6151–6167, https://doi.org/10.5194/acp-10-6151-2010, 2010.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: Inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, https://doi.org/10.5194/bg-9-457-2012, 2012.

Harris, I. C.: CRU JRA v1. 1: A forcings dataset of gridded land surface blend of Climatic Research Unit (CRU) and Japanese reanalysis (JRA) data; Jan. 1901-Dec. 2017, University of East Anglia Climatic Research Unit, Centre for Environmental Data Analysis, https://doi.org/10.5285/13f3635174794bb98cf8ac4b0ee8f4ed, 2019.

Heskel, M. A., O'Sullivan, O. S., Reich, P. B., Tjoelker, M. G., Weerasinghe, L. K., Penillard, A., Egerton, J. J., Creek, D., Bloomfield, K. J., Xiang, J., Sinca, F., Stangl, Z. R., Martinez-De La Torre, A., Griffin, K. L., Huntingford, C., Hurry, V., Meir, P., Turnbull, M. H., and Atkin, O. K.: Convergence in the temperature response of leaf respiration across biomes and plant functional types, Proc. Natl. Acad. Sci. U. S. A., 113, 3832–3837, https://doi.org/10.1073/pnas.1520282113, 2016.

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., Dlugokencky, E., Tans, P. P., Shiga, Y. P., Mountain, M., Nehrkorn, T., Montzka, S. A., McKain, K., Kofler, J., Trudeau, M., Michel, S. E., Biraud, S. C., Fischer, M. L., Worthy, D. E., Vaughn, B. H., White, J. W., Yadav, V., Basu, S., and Van Der Velde, I. R.: Enhanced North American carbon uptake associated with El Niño, Sci. Adv., 5, 1–11, https://doi.org/10.1126/sciadv.aaw0076, 2019.

Huntzinger, D. N., Gourdji, S. M., Mueller, K. L., and Michalak, A. M.: The utility of continuous atmospheric measurements for identifying biospheric CO_2 flux variability, J. Geophys. Res. Atmos., 116, 1–14, https://doi.org/10.1029/2010JD015048, 2011.

Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., Schaefer, K., Wei, Y., Cook, R. B., Fisher, J. B., Hayes, D., Huang, M., Ito, A., Jain, A. K., Lei, H., Lu, C., Maignan, F., Mao, J., Parazoo, N., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tian, H., Wang, W., Zeng, N., and Zhao, F.: Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions, Sci. Rep., 7, 1–8, https://doi.org/10.1038/s41598-017-03818-2, 2017.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y. P.,

Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO_2 sink changes to temperature, Nature, 541, 516–520, https://doi.org/10.1038/nature20780, http://dx.doi.org/10.1038/nature20780, 2017.

King, A. W., Andres, R. J., Davis, K. J., Hafer, M., Hayes, D. J., Huntzinger, D. N., De Jong, B., Kurz, W. A., McGuire, A. D., Vargas, R., Wei, Y., West, T. O., and Woodall, C. W.: North America's net terrestrial CO_2 exchange with the atmosphere 1990-2009, Biogeosciences, 12, 399–414, https://doi.org/10.5194/bg-12-399-2015, 2015.

Kitanidis, P. K.: Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis, Water Resour. Res., 22, 499–507, https://doi.org/10.1029/WR022i004p00499, 1986.

Kuze, A., Suto, H., Shiomi, K., Nakajima, M., and Hamazaki, T.: On-orbit performance and level 1 data processing of TANSO-FTS and CAI on GOSAT, Sensors, Syst. Next-Generation Satell. XIII, 7474, 74 740I, https://doi.org/10.1117/12.830152, 2009.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurtt, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lima, I., Lombardozzi, D., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., Nojiri, Y., Padín, X. A., Peregon, A., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wiltshire, A. J., Zaehle, S., and Zhu, D.: Global Carbon Budget 2018 (pre-print), Earth Syst. Sci. Data Discuss., pre print, 1–54, https://www.earth-syst-sci-data.net/10/2141/2018/essd-10-2141-2018.pdf, 2018.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y., O'Dell, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., and Eldering, A.: Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño, Science (80-.)., 358, https://doi.org/10.1126/science.aam5690, 2017.

Luus, K. A., Commane, R., Parazoo, N. C., Benmergui, J., Euskirchen, E. S., Frankenberg, C., Joiner, J., Lindaas, J., Miller, C. E., Oechel, W. C., Zona, D., Wofsy, S., and Lin, J. C.: Tundra photosynthesis captured by satellite-observed solar-induced chlorophyll fluorescence, Geophys. Res. Lett., 44, 1564–1573, https://doi.org/10.1002/2016GL070842, 2017.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochem. Cycles, 22, https://doi.org/10.1029/2006GB002735, 2008.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res. D Atmos., 109, 1–19, https://doi.org/10.1029/2003JD004422, 2004.

Miller, S. M. and Michalak, A. M.: The impact of improved satellite retrievals on estimates of biospheric carbon balance, Atmos. Chem. Phys., 20, 323–331, https://doi.org/10.5194/acp-20-323-2020, 2020.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., et al.: Anthropogenic emissions of methane in the United States, Proceedings of the National Academy of Sciences, 110, 20 018–20 022, https://doi.org/10.1073/pnas.1314392110, 2013.

Miller, S. M., Michalak, A. M., Yadav, V., and Tadié, J. M.: Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite, Atmos. Chem. Phys., 18, 6785–6799, https://doi.org/10.5194/acp-18-6785-2018, 2018.

Miller, S. M., Saibaba, A. K., Trudeau, M. E., Mountain, M. E., and Andrews, A. E.: Geostatistical inverse modeling with very large datasets: An example from the Orbiting Carbon Observatory 2 (OCO-2) satellite, Geosci. Model Dev., 13, 1771–1785, https://doi.org/10.5194/gmd-13-1771-2020, 2020.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, J. Geophys. Res. Atmos., 113, 1–15, https://doi.org/10.1029/2007JD009734, 2008.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cycles, 24, https://doi.org/10.1029/2009GB003642, 2010.

Nakajima, M., Kuze, A., and Suto, H.: The current status of GOSAT and the concept of GOSAT-2, International Society for Optics and Photonics, 8533, 853 306, https://doi.org/10.1117/12.974954, 2012.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): A global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth Syst. Sci. Data, 10, 87–107, https://doi.org/10.5194/essd-10-87-2018, 2018.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M., Feng, L., Palmer, D. P. I., Dubey, M., García, O. E., Griffith, D. W., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., et al.: Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, BioScience, 51, 933–938, https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

Palmer, P. I., Feng, L., Baker, D., Chevallier, F., Bösch, H., and Somkuti, P.: Net carbon emissions from African biosphere dominate pantropical atmospheric CO_2 signal, Nat. Commun., 10, 3344, https://doi.org/10.1038/s41467-019-11097-w, http://www.nature.com/articles/s41467-019-11097-w, 2019.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrol. Earth Syst. Sci., 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., and Wang, X.: Global Biogeochemical Cycles simulated by terrestrial ecosystem models, Global Biogeochem. Cycles, 29, 46–64, https://doi.org/10.1002/2014GB004931.Received, 2014.

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., Ma, Y., Myneni, R. B., Poulter, B., Sun, Z., Wang, T., Viovy, N., Zaehle, S., and Zeng, N.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, Glob. Chang. Biol., 19, 2117–2132, https://doi.org/10.1111/gcb.12187, 2013.

Piao, S., Liu, Z., Wang, T., Peng, S., Ciais, P., Huang, M., Ahlstrom, A., Burkhart, J. F., Chevallier, F., Janssens, I. A., Jeong, S. J., Lin, X., Mao, J., Miller, J., Mohammat, A., Myneni, R. B., Peñuelas, J., Shi, X., Stohl, A., Yao, Y., Zhu, Z., and Tans, P. P.: Weakening temperature control on the interannual variations of spring carbon uptake across northern lands, Nat. Clim. Chang., 7, 359–363, https://doi.org/10.1038/nclimate3277, 2017.

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., Friedlingstein, P., and Sitch, S.: Interannual variation of terrestrial carbon cycle: Issues and perspectives, Glob. Chang. Biol., 26, 300–318, https://doi.org/10.1111/gcb.14884, 2020.

Poulter, B., Heyder, U., and Cramer, W.: Modeling the sensitivity of the seasonal cycle of GPP to dynamic LAI and soil depths in tropical rainforests, Ecosystems, 12, 517–533, https://doi.org/10.1007/s10021-009-9238-4, 2009.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., et al.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, Nature, 509, 600–603, https://doi.org/10.1038/nature13376, 2014.

Raich, J. W., Rastetter, E. B., Melillo, J. M., Kicklighter, D. W., Steudler, P. A., Peterson, J., Grace, A. L., Iii, B. M., and Vörösmarty, C. J.: Potential Net Primary Productivity in South America : Application of a Global Model, Ecol. Appl., 1, 399–429, 1991.

Ramsey, F. and Schafer, D.: The statistical sleuth: a course in methods of data analysis, Cengage Learning, 2012.

Randerson, J., Van der Werf, G., Giglio, L., Collatz, G., and Kasibhatla, P.: Global Fire Emissions Database, Version 4.1 (GFEDv4), ORNL DAAC, Oak Ridge, Tennessee, USA, https://doi.org/0.3334/ORNLDAAC/1293, 2018.

Ratner, B.: Statistical and Machine-Learning Data Mining:: Techniques for Better Predictive Modeling and Analysis of Big Data, CRC Press, 2012.

Restrepo-Coupe, N., Levine, N. M., Christoffersen, B. O., Albert, L. P., Wu, J., Costa, M. H., Galbraith, D., Imbuzeiro, H., Martins, G., da Araujo, A. C., Malhi, Y. S., Zeng, X., Moorcroft, P., and Saleska, S. R.: Do dynamic global vegetation models capture the seasonality of carbon fluxes in the Amazon basin? A data-model intercomparison, Glob. Chang. Biol., 23, 191–208, https://doi.org/10.1111/gcb.13442, 2016.

Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to interannual climatic variations? A quantification based on atmospheric CO₂ data, Biogeosciences Discuss., pp. 1–22, https://doi.org/10.5194/bg-2018-34, 2018.

Schwarz, G. et al.: Estimating the dimension of a model, The annals of statistics, 6, 461–464, 1978.

Shiga, Y. P., Michalak, A. M., Fang, Y., Schaefer, K., Andrews, A. E., Huntzinger, D. H., Schwalm, C. R., Thoning, K., and Wei, Y.: Forests dominate the interannual variability of the North American carbon sink, Environ. Res. Lett., 13, https://doi.org/10.1088/1748-9326/aad505, 2018.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quéré, C., Smith, B., Zhu, Z., and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon dioxide, Biogeosciences, 12, 653–679, https://doi.org/10.5194/bg-12-653-2015, 2015.

Wang, K., Wang, Y., Wang, X., He, Y., Li, X., Keeling, R. F., Ciais, P., Heimann, M., Peng, S., Chevallier, F., Friedlingstein, P., Sitch, S., Buermann, W., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E., Poulter, B., Vuichard, N., Wiltshire, A., Zeng, N., Zhu, D., and Piao, S.: Causes of slowing-down seasonal CO_2 amplitude at Mauna Loa, Glob. Chang. Biol., 26, 4462–4477, https://doi.org/10.1111/gcb.15162, 2020.

Wang, X., Piao, S., Ciais, P., Friedlingstein, P., Myneni, R. B., Cox, P., Heimann, M., Miller, J., Peng, S., Wang, T., Yang, H., and Chen, A.: A two-fold increase of carbon cycle sensitivity to tropical temperature variations, Nature, 506, 212–215, https://doi.org/10.1038/nature12915, 2014.

Wofsy, S., Afshar, S., Allen, H., Apel, E., Asher, E., Barletta, B., Bent, J., Bian, H., Biggs, B., Blake, D., et al.: ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, ORNL DAAC, Oak Ridge, Tennessee, USA, https://doi.org/10.3334/ORNLDAAC/1581, 2018.

Wunch, D., Toon, G. C., Blavier, J. F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W., Sherlock, V., and Wennberg, P. O.: The total carbon column observing network, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655–2671, https://doi.org/10.5194/bg-7-2655-2010, 2010.

Yang, D., Liu, Y., Cai, Z., Chen, X., Yao, L., and Lu, D.: First Global Carbon Dioxide Maps Produced from TanSat Measurements, Adv. Atmos. Sci., 35, 621–623, https://doi.org/10.1007/s00376-018-7312-6, 2018.

Zucchini, W.: An introduction to model selection, J. Math. Psychol., 44, 41–61, https://doi.org/10.1006/jmps.1999.1276, 2000.

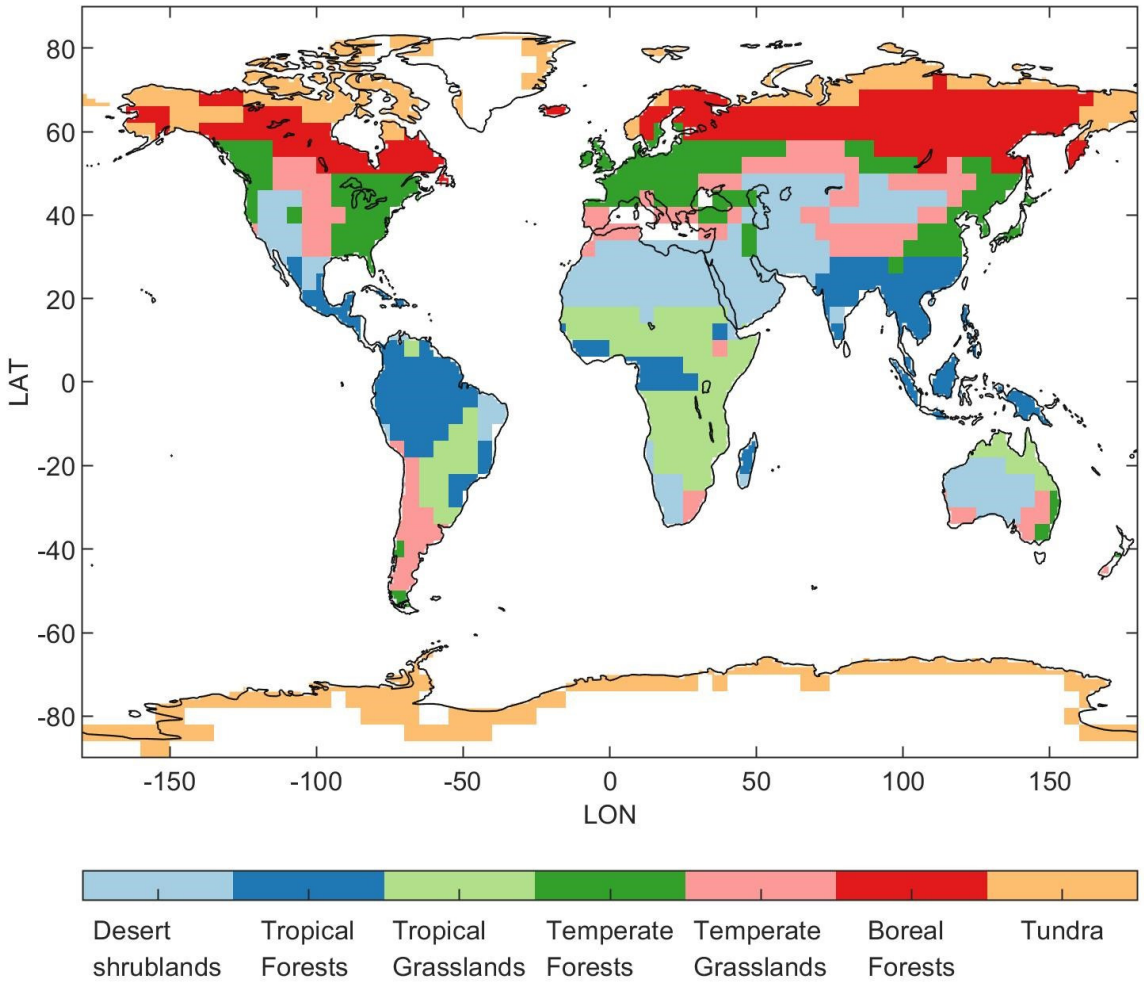

Figure 1. The seven biome-based regions aggregated from a world biome map in Olson et al. (2001).

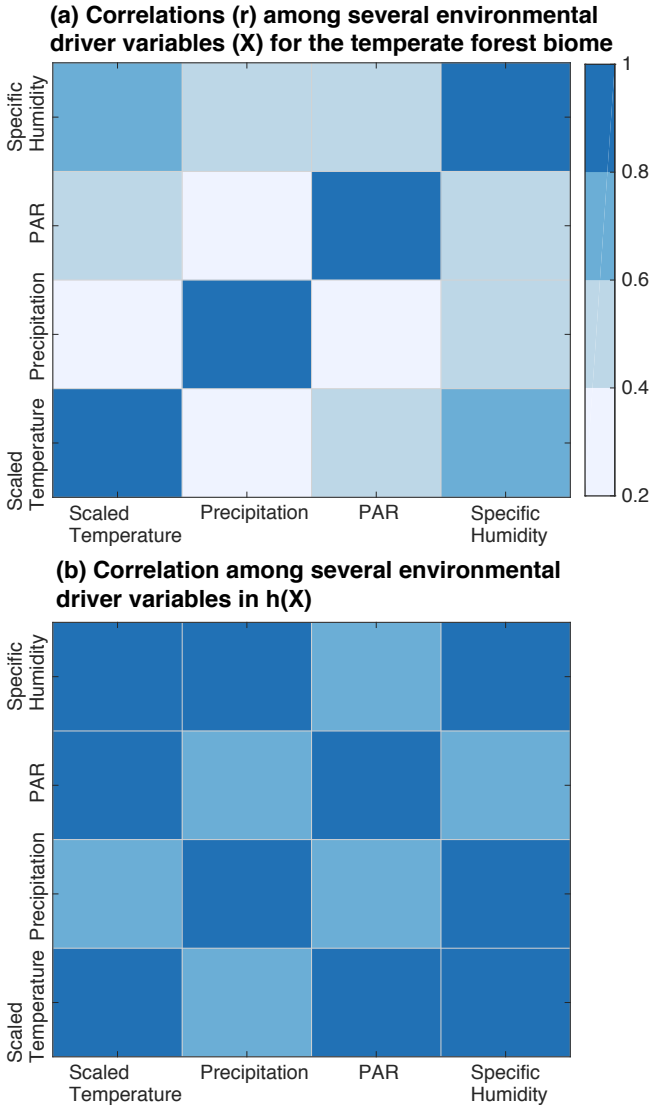

Figure 2. Correlation coefficients (r) between environmental drivers over temperate forests biome in year 2017 in $\mathbf{X}$ (panel a) and $h(\mathbf{X})$ (panel b). We find that the correlation between environmental drivers ($\mathbf{X}$) are generally low (a), e.g., PAR and scaled temperature, precipitation and specific humidity; however, when these environmental drivers are passed through the transport model $h()$ and interpolated to the locations of OCO-2 observations, the correlation between these drivers become much stronger (b), indicating high collinearity.

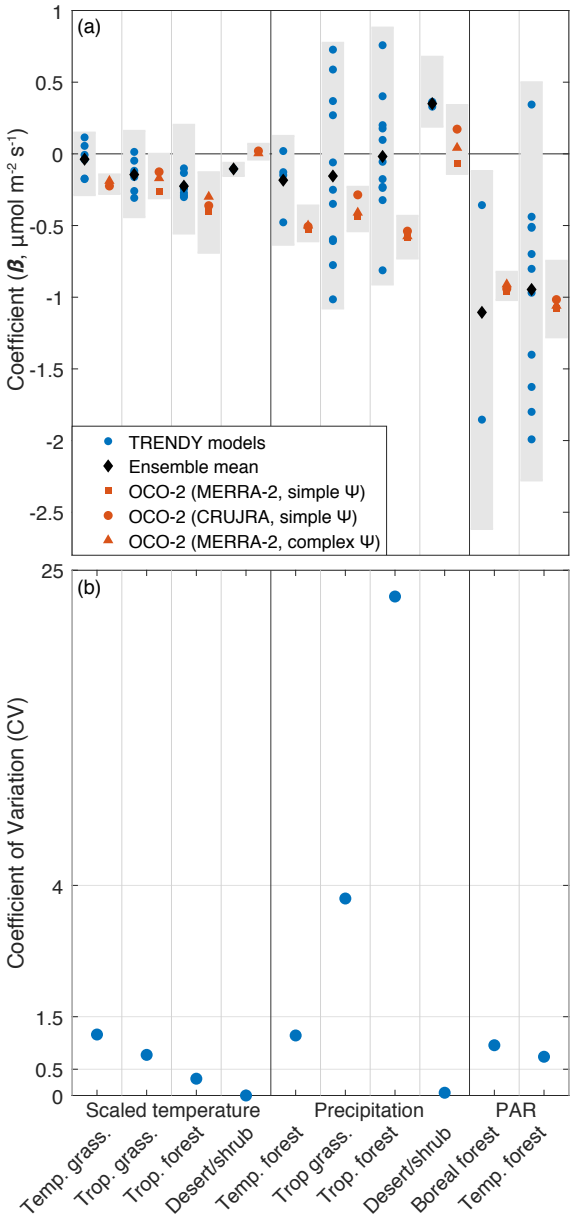

Figure 3. Estimated coefficients (β) from the TRENDY models (blue), from the ensemble mean of the TRENDY models (black), and from the analysis using OCO-2 (red). Each blue or red dot indicates the mean value across all four years of the study period. Gray bars indicate the full range of uncertainties in the coefficients. To construct these gray bars, we calculate the uncertainties in the coefficients estimated for each individual TBM (or for the real OCO-2 data) using Eq. 4. They gray bars encapsulate all of the uncertainty bounds from all of these individual model calculations. Furthermore, the analysis of OCO-2 includes simulations using MERRA-2 meteorology with a simple formulation of Ψ (red square), using CRUJRA meteorology and a simple formulation of Ψ (red dot), and using MERRA-2 and a complex formulation of Ψ (the same used in the GIM, red triangle). The coefficients from the analysis using OCO-2 (red) are broadly within the range of the estimates in TBMs (blue). We further calculate the coefficient of variation (CV) of coefficients for each environmental drivers within the TBMs (b), and we find that the largest CV are from the coefficients for precipitation.

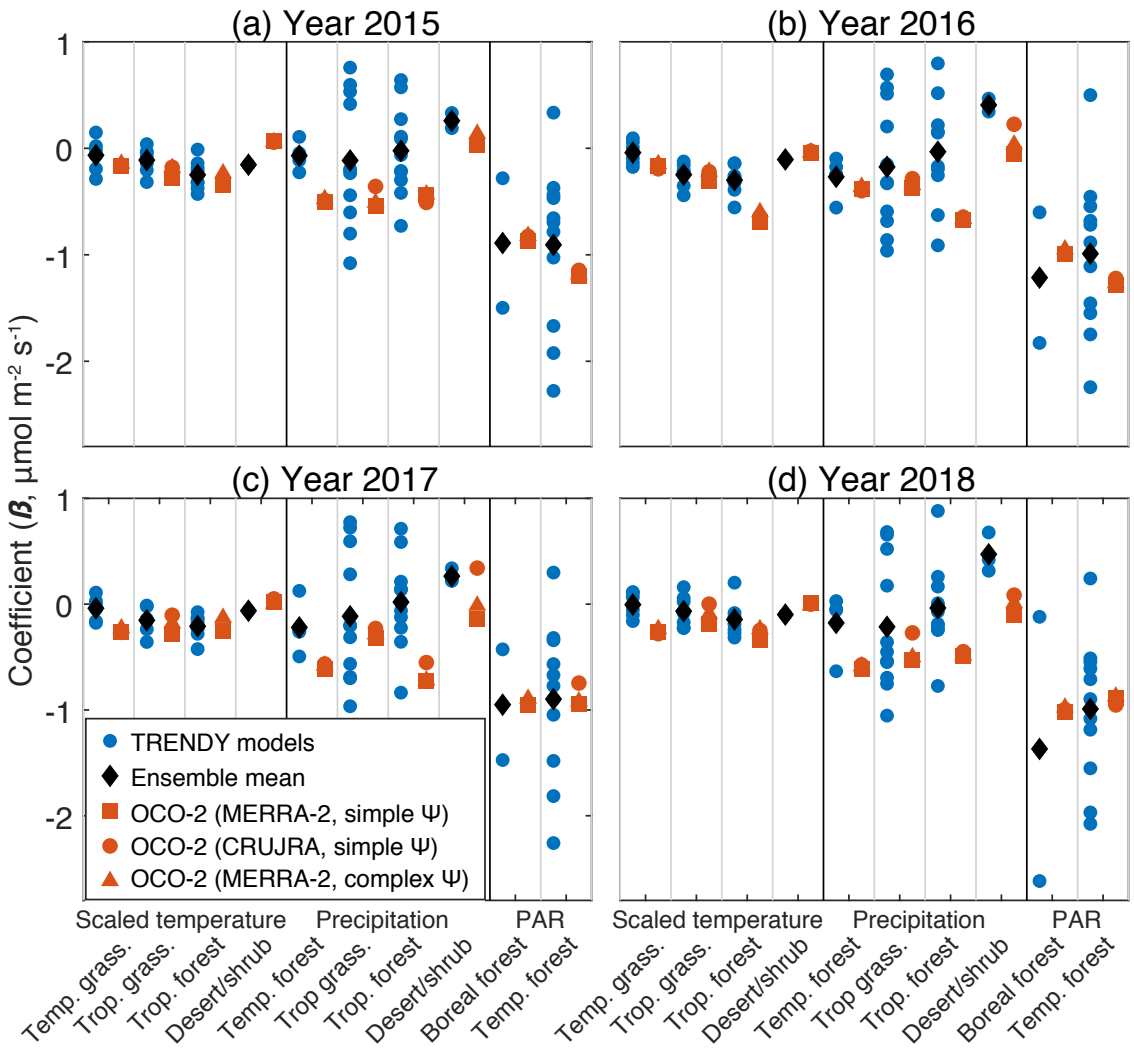

Figure 4. This figure is similar to Fig. 3 but shows results for individual years. There are no noticeable shifts in the coefficient estimates between El Niño (2015 – 2016; panels a-b) and non-El Niño years (2017 – 2018; panels c-d) from the analysis using OCO-2 (red). Some individual TBMs show differences of up to 50% in the estimated coefficient among years, though many individual TBMs do not.

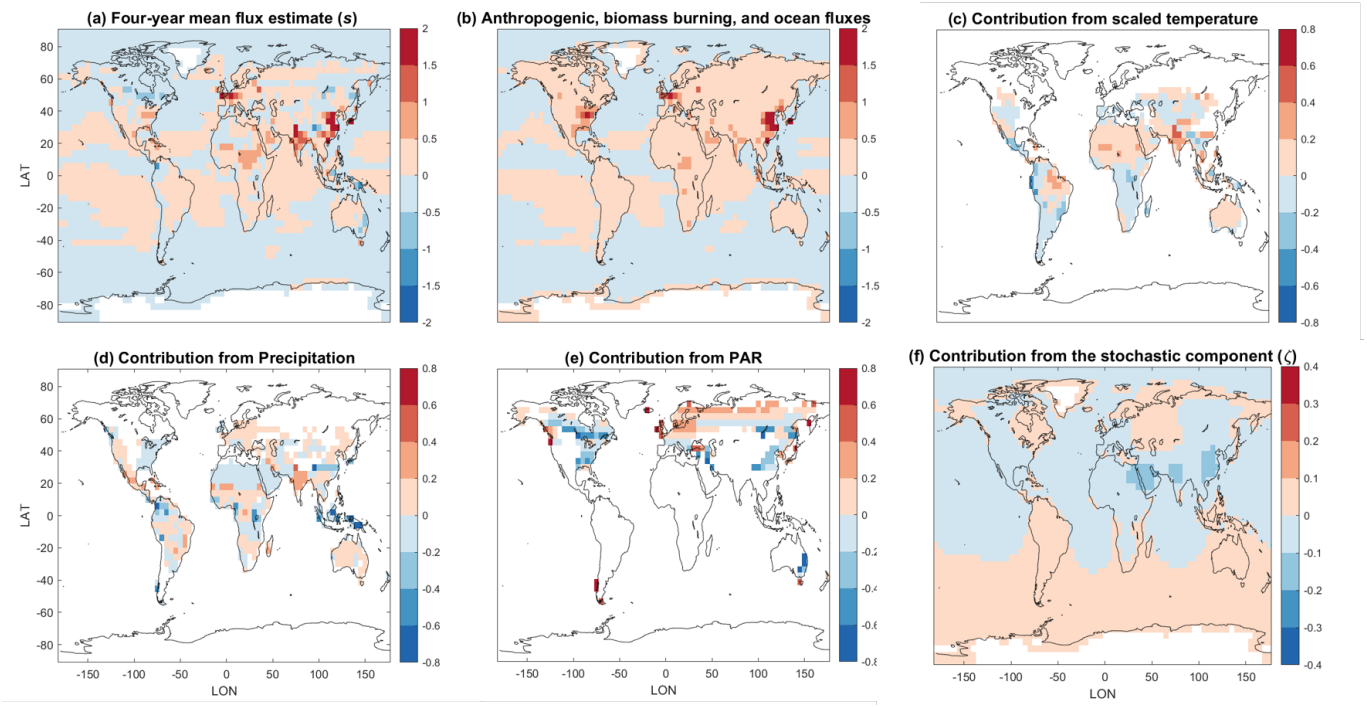

Figure 5. The contribution of different environmental driver datasets to the flux estimate from the GIM. Panel (a) displays the four-year mean flux estimate (including both the regression and stochastic components of the flux estimate; units of μmol m^{-2} s^{-1}) and panel (b) the contribution from anthropogenic, biomass burning, and ocean fluxes. Contributions from different environmental drivers, including scaled temperature (c), precipitation (d), and PAR (e), describe most of spatiotemporal variability in terrestrial biospheric CO_2 fluxes, whereas the stochastic components (ζ), panel f) only account for a small portion of flux variability. Note that the inverse modeling results shown in this figure use environmental driver data from MERRA-2. Also note the color bars used in panels (a-b), panels (c-e), and panel (f) are different. White colors in panels (c-e) indicates that not all environmental drivers are selected in all biomes.

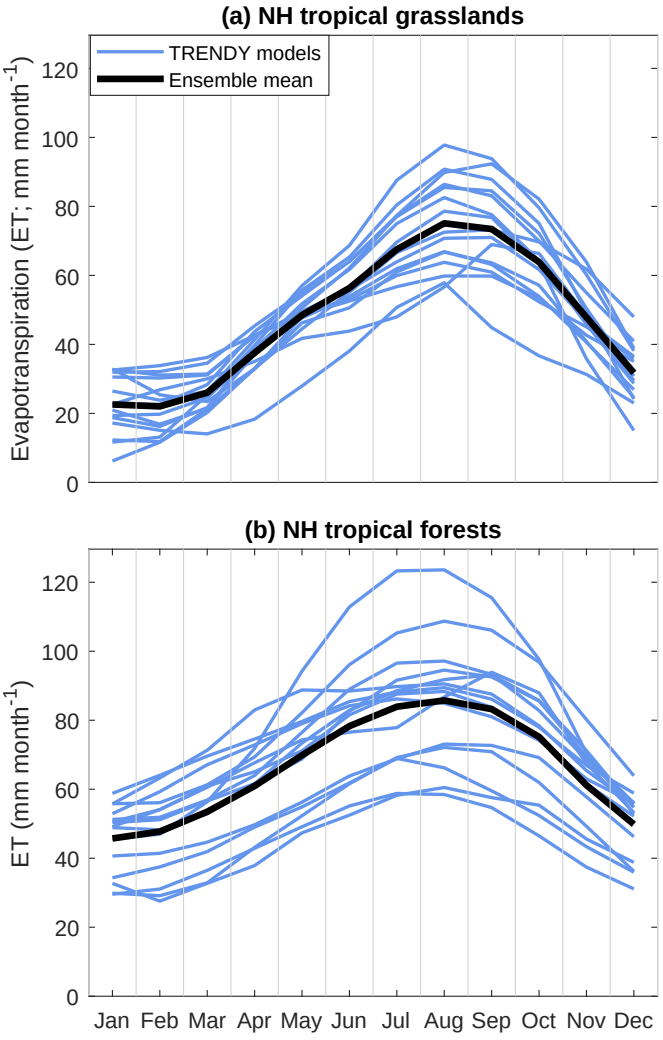

Figure 6. Four-year averaged evapotranspiration (ET) estimates from a suite of 15 TBMs (blue) and from the ensemble mean (black), for North Hemispheric tropical grasslands (a) and for North Hemispheric tropical forests (b). Annual ET show large differences in magnitude across the TBMs for both tropical biomes.