

We thank Abhishek Chatterjee and an anonymous reviewer for their constructive comments. We have addressed the comments of the reviewers below. Reviewer comments are shown in bold and changes to the manuscript are underlined. Line numbers refer to the marked changes manuscript.

Both reviewers found that the manuscript contained excessive detail that obscured the main results of the study. We have gone through the manuscript to remove excessive detail and present the results more concisely. In particular, we have shortened the introduction, removed any discussion from the results section, and shortened the discussion section. Some of the analysis that is complementary to the main results has been moved to the supplementary analysis. These steps have reduced the length of the manuscript by about 10 pages.

Response to reviewer 1:

### **Holistic suggestions**

**The authors have written an interesting paper that is very relevant to current inverse modeling efforts using satellite data. Overall, I think the science in this paper is very sound. The authors have been very thorough and thoughtful. With that said, the manuscript contains enormous detail and discussion. I worry that the main takeaway messages can get lost in this detail. I think the main science question posed in this paper could have a relatively straightforward answer; I think one could answer that question in a more succinct way that brings to the forefront the unique results of the paper. Here are a few ways that one could shorten the paper: (1) Some text in the introduction seems better-suited for the methods (see below), and I think the intro could end with a more focused, succinct list of the study goals. (2) Some of the details on the proxies and the flux products could go into a Supplement, though I think it could go either way. (3) I would consider merging the results and discussion into a single section. These two sections can feel repetitive in places. (4) I think the Conclusions could be half the current length and focus on bringing the analysis together into snappy, punchy take-away messages. These are just a few ideas, and I think the authors should make whatever changes feel most appropriate to them. I want to be sure this article gets the attention it deserves when published and ensure that the authors don't lose readers in the detail. Below are several detailed suggestions.**

We have gone through the manuscript and tried to remove any unnecessary text and reorganized some of the text to improve the clarity of the manuscript.

### **Abstract**

- **You may want to either define FLUXCOM or use a non-technical term here.**

As far as we are aware, FLUXCOM is not an acronym but the name of a flux product. We have added the URL for the datasets immediately after the first occurrence of "FLUXCOM".

- **What is R<sub>NI</sub>NO<sub>3.4</sub>? Again, you might want to either define this term or use a less technical term.**

In the abstract, we have changed the text to state the “statistical significance” (P<0.05) rather than using this term, as the reader will likely be familiar with this term.

- **Line 14: Could you define “regional scales” more precisely? I.e., what size regions are you referring to here?**

Clarified to “large sub-continental regions”

- **Line 14: It could be helpful to give the reader a hint on what aspects of the configuration you’re referring to here.**

This statement was removed from the abstract.

## **Introduction and methods**

- **pg. 2, Line 23, “We also examine the posterior fluxes from two publicly available flux inversion estimates....”: Can you provide a bit more detail here. In what way do you examine these fluxes, and how does this examination contribute to the study?**

This sentence has been re-worded:

P3L1-2: “NEE anomalies produced by the GEOS-Chem flux inversions are contrasted with two independent publicly available flux inversion estimates.”

- **I would write out abbreviations like MERRA-2, FLUXNET, and FLUXCOM the first time you use them.**

Done for MERRA2. FLUXNET and FLUXCOM are not acronyms, as far as we are aware. We have added the URL for the datasets immediately after the first occurrence of FLUXNET and FLUXCOM.

- **Pg. 2, line 32 to pg. 4, line 9: The introduction is relatively long. You could condense or eliminate some of this information to produce a shorter, punchier introduction.**

We have tried to condense the introduction, and have shortened it.

- **Pg. 4, line 3 to line 15: This material could be a better fit for the methods sections.**

We have removed this material from the introduction.

- **Pg. 4, lines 16-27: Again, I think you could shorten the descriptions here to make the main objectives more concise and punchy. There’s a lot of detail to digest in this paragraph. It could be helpful to give a more succinct overview of the paper and leave some of the details for later.**

We have tried to do this; the introduction has been restructured for this purpose.

• **Pg. 7, line 3: I recommend experimenting with covariance matrices that include off-diagonal elements. I worry that you could overestimate the information content of the satellite observations by using a diagonal observational covariance matrix. Existing studies suggest that GOSAT retrievals have spatially and temporally coherent errors. My guess is that the estimated fluxes would be less detailed or less informative if you included off-diagonal elements in the observational covariance matrix. The overall goal of this study is to estimate the robustness of the flux constraint using GOSAT observations, and I think that the structure of the covariance matrices could have an important impact on these conclusions.**

We agree that there are likely spatially and temporally coherent errors in the ACOS GOSAT measurements. However, implementing off-diagonal elements in our inversion set-up is not straight forward, and would be a study on its own. Thus, this is beyond the scope of this study. We have added a statement to the conclusions encouraging future research in this field:

P24L20-22: “Although not addressed in this study, correlated errors between GOSAT observations may introduce structures in the posterior NEE estimates, thus we recommend future work address the possibility of prescribing non-diagonal in the observational error covariance matrix.”

## **Results and conclusions**

• **Pg. 10, line 26, “As scales decrease....”:** It could be clearer to rephrase to “As the size of the regions decrease....”.

Changed to “At smaller spatial scales”

• **Pg. 10, line 32: I would be clearer about what you mean by “regional and larger scales.”** Different readers might have different concepts of what this phrase means.

We have specified “sub-continental regions” whenever we mention “regional” scales

• **In general, there’s a lot of detail in Sect. 3. I think you could shorten this section and/or potentially move some of the text to an SI without losing key information. An upside of being so thorough is that you won’t omit anything from the text. However, a danger is that the key messages of the paper can become buried in all of the detailed descriptions.**

We have removed a substantial amount of text from Section 3, and ensured that the remaining text only described the results without giving a discussion.

• **Sect. 3.1.1 title: I think it would be more appealing to a reader to use a descriptive title here instead of using a relatively technical abbreviation in the header title. Someone who is scanning through the paper wouldn’t know what GC<sub>2x2.5-200%</sub> is or what the scientific importance of this section is.**

This section has been moved to supplementary materials and renamed to “Detailed analysis of tropical NEE anomalies and ENSO”.

- **Pg. 14, line 18: Consider starting this section with a stronger topic sentence that helps guide the reader through the material. For example, you could state the main take-away message of the OSSEs in the tropics and then explain the results in subsequent sentences.**

Changed topic sentence:

P15L18: “Strong correlations are obtained between the posterior and true anomalies for all OSSEs on sub-continental regional and larger scales, suggesting that sub-continental regions are the minimum scales that can be constrained by GOSAT measurements.”

- **Sect. 3.3.1: I think you could add a few more sentences somewhere in this section explaining what the implications of the OSSE’s are for your real data results. What have you learned from the OSSE’s, and what are the implications for using GOSAT to constrain flux anomalies?**

Added to beginning of the section:

P15L10:” Strong correlations between the GOSAT flux inversions and proxies/FLUXCOM provide evidence that the GOSAT flux inversions give realistic constraints on NEE. However, the absence of strong correlations does not imply that the GOSAT flux inversions are not constraining IAV as there could be other causes (such as lagged effects within ecosystems) that can explain the absence of correlations. Therefore, to investigate the minimum spatial scales that can be constrained by GOSAT observations, we performed a series of OSSE experiments. In these experiments pseudo-observations were assimilated from a GEOS-Chem forward model run which had JULES NEE fluxes prescribed.”

- **Pg. 20, line 1: I think it is a bit too informal to start a paragraph with a question. You might want to rephrase with a guiding topic sentence (e.g., one that states the main result or take-away of the paragraph).**

Removed this sentence, and tried to make this section more concise.

- **You use the term “flux anomalies” a lot throughout the paper. I would make sure it’s very clear to the reader what type of anomalies you are referring to (e.g., over what time periods). I know that you specify this information in the Methods, but it doesn’t hurt to use very clear and unambiguous terminology whenever possible.**

We have edited the manuscript throughout to clarify quantity and time interval.

- **Pg. 21, line 19 “significant impact”: Is this result statistically significant? I would reserve the word “significant” for instances when you explicitly tested for statistical significance. Otherwise, you could unintentionally confuse a reader.**

Removed.

Response to Reviewer 2:

**This is a highly relevant and high-quality study that makes use of the ever-increasing time record of satellite retrievals (in this case JAXA's GOSAT), and demonstrates their value in providing the community with new insights into carbon cycle science. While most inversion studies tend to diagnose carbon fluxes, this study makes a valuable attempt at attribution by relating the carbon flux anomalies with various auxiliary environmental variables (called proxies here). The manuscript is well-written and the figures and text are of high-quality. Before recommending final publication though, I would highly advise the authors and the ACP editorial team to make note of the following comments and suggestions:**

**Major Comments:**

- (1) In many ways, the manuscript reads like a dissertation chapter. There are too many details that are unnecessary for a manuscript as it outlines all the sensitivity tests and checks that were done. I would highly recommend that the authors think about what the main takeaway messages are, and accordingly cull the text from Sections 3, 4 and 5 to focus on those messages. There are some strong points made in the paper, for example - the need to use prior flux estimates that have a realistic seasonal cycle and IAV. But more often than not, these types of important messages are lost in the text. Even though the manuscript is well-written, it is cumbersome to read. Some of the sensitivity tests could be moved to Supplementary Information. Also, see comments #2 & #4 below on which text are most suited for this.**

We have gone through the manuscript and tried to streamline to highlight the main points. The manuscript is now substantially shorter.

- (2) Is SIF really needed? As the authors rightly point out in multiple places (Page 3, Lines 24-29, Page 11, discussion in Lines 17-30, Page 16, Lines 30-32), SIF is a good proxy for GPP. Its relationship to NEE is complicated by the role of R. If the authors were estimating GPP and R separately, then SIF would have been a valuable dataset to correlate against. But given that the authors estimate NEE, it is impractical to assume SIF will provide any new information or help constrain the scales at which GOSAT can inform NEE fluxes. Results presented in Figure 2 and Table 2 further show the redundancy of SIF for this paper. In fact, it is not surprising at all that SIF is well correlated with CASA GFED ( a model based on light-use efficiency) and consequently CT-2016 (that uses CASA-GFED as its prior). Finally, Section 4.1.2 doesn't add any new information to the study at all. By taking out SIF, the authors could shorten the paper and make it more focused.**

We prefer to retain SIF in the manuscript, as this is the only proxy that is a direct measure of ecosystem functioning (even if it is a measure of GPP). Furthermore, we do find substantial correlations between SIF anomalies and NEE anomalies in the Northern extra

tropics. Similarly, we see that large anomalies and heatwaves are shown to have a large impact on ecosystem functioning in Figure 3.

- (3) Follow-on to the previous comment - if the authors were interested in using a vegetation index, they may want to look at something more direct like LAI. Or even indices like NDVI, EVI, fPAR x PAR may act as robust proxy capturing the interaction between vegetation and radiation.**

There are many interesting potential “proxies” for ecosystem functioning that we could examine. However, we think that the inclusion of additional proxies would distract from our main point that we can see strong correlations with the proxies chosen for this study on a variety of scales.

- (4) Should the OSSEs be moved to the Supplementary Material? While the scientific rationale for doing the OSSEs make sense, they add a new set of results that are frankly not necessary. The real data inversions already cover all the conclusions from the OSSEs. I understand that the OSSEs may have been done initially to ensure that the GEOS-Chem 4DVAR system works, before the authors move to a real-data inversion. But given the more interesting and thought provoking results with the real data, Section 3.3 seems superfluous.**

We prefer to retain the OSSEs in the main text. This is because the OSSEs are required to estimate the minimum scales constrained by the GOSAT measurements and estimate the magnitude of posterior IAV relative to true IAV. Neither of these aspects of the GOSAT flux inversions can be examined in the real data experiments.

- (5) The attempt to isolate anomalies specific to each continent (Equation 2) is unclear. I do not follow how the standard deviation of the tropical and the continental anomaly are calculated in the first place. Do the authors do some type of Monte Carlo simulation (i.e., an ensemble of fluxes and anomalies) to calculate STD(ANOM).**

It is calculated as the standard deviation of the monthly anomalies. We have decided to move this discussion to the supplementary materials as the main conclusions of the analysis could be made without this analysis. We have revised the sentence in the supplementary materials to state “standard deviation of monthly anomalies”.

- (6) A critical component that is missing from the study is the calculation of time-lagged correlations. The impact of ENSO events or droughts are not immediate (i.e., within a month) on fluxes. There is a spatiotemporal cascade of this impact across different tropical and Northern extratropical regions. In fact the authors do acknowledge this briefly (Page 16, Lines 17-20). Also see Rayner et al. 1999 (GRL, Vol. 26(4)). My hypothesis is that both the R2 and the significance of several of the flux estimates will change when the authors use time-lagged correlation, especially with the Nino3.4 index. Can the authors comment further on why they didn't pursue? A figure or two in Supplementary Materials may help.**

We agree that there are substantial lagged effects between climate anomalies and NEE anomalies, and that these lagged effects are important for understanding the carbon cycle. However, we have intentionally avoided discussing lagged effects in this manuscript to simplify the analysis. As pointed out by the reviewer, we do acknowledge that these lagged effects exist in the manuscript. We hope that these effects will be investigated further in future studies, and have suggested this is the revised conclusions:

P24L26-31:“The results of this study suggest that GOSAT measurements provide a useful constraint on IAV in the carbon cycle. Further study of the relationship between GOSAT-constrained NEE and environmental variables is merited given results discussed here. In particular, the mechanisms driving these co-variations should be further investigated. Lagged relationships between GOSAT-constrained NEE and environmental variables should also be investigated. Future research could also investigate differences in IAV between GOSAT-constrained NEE and that produced by TBMs. Given the better agreement with the proxies, GOSAT-constrained NEE IAV may provide a tool for evaluating the TBM-simulated NEE IAV in the future.”

**(7) Finally, why isn't there a run at 2 x 2.5 with prescribed IAV in the prior flux? If anything, I would want to see how an inversion run at 2 x 2.5 with 200% uncertainty applied to prior fluxes and with prior NEE IAV performs. The results may really enforce the message in Section 4.3.3 about the realism of IAV in prior fluxes and the important role it plays on the posterior IAV that is recovered during the inversion.**

This was not performed because of the computation requirements. Performing 4DVar at 2x2.5 spatial resolution with the GEOS-Chem is extremely computationally expensive, thus we tried to perform all of the sensitivity tests at 4x5 degrees.

#### **Minor Comments:**

**(1) Page 5, Section 2.1 - What is the spatial and temporal resolution at which FLUXCOM products are available? Are there potentially issues with scale mismatch that the reader should be aware of?**

The FLUXCOM products are generated at 0.5x0.5 degrees resolution. This has been added to the text.

P5L18: “FLUXCOM remote sensing and meteorological data (RS+METEO) products are generated at 0.5 x 0.5 spatial resolution using upscaling approaches based on machine learning methods that integrate FLUXNET site level observations, satellite remote sensing, and meteorological data”

**(2) Page 7, Lines 30-33 -The current nomenclature for the different runs is extremely complex. I do not know if there is a way of making the names more intuitive and/or easy to follow. But something the authors can think about a bit more.**

We tried a number of different acronyms, but found the current set-up to be the clearest. We hope that table 1 will provide a useful tool to remind the reader of the set-up.

**(3) Page 8, Line 3 - Did GOSAT ACOS v3.5 had warn levels? Or just quality flags? To the best of my knowledge warn levels weren't implemented till v7.3. Kindly check.**

Warn levels were implemented in ACOS 3.5. This was the first ACOS GOSAT product to have them.

**(4) Page 9, Equation 1 - It is worth making the reader aware of the drawbacks of using a short period to define the climatology.**

We have added the following text.

P10L9-11: "It is worth noting that four years is a relatively short period to define a climatology, and some modes of climate variability occur on longer timescales. Ideally, a longer time period would be used to calculate a climatology, but we are limited by the availability of GOSAT data in this study."

**(5) Section 4 Discussion topics - what about the discussion for agreement between flux anomalies and Nino 3.4 index? That discussion seems to have gone missing.**

This has been added as Section 4.2.



# On what scales can GOSAT flux inversions constrain anomalies in terrestrial ecosystems?

Brendan Byrne<sup>1,\*</sup>, Dylan B. A. Jones<sup>1,2</sup>, Kimberly Strong<sup>1</sup>, Saroja M. Polavarapu<sup>3</sup>, Anna B. Harper<sup>4</sup>, David F. Baker<sup>5,6</sup>, and Shamil Maksyutov<sup>7</sup>

<sup>1</sup>Department of Physics, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, USA

<sup>3</sup>Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada

<sup>4</sup>College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, UK

<sup>5</sup>NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, CO, USA

<sup>6</sup>Cooperative Institute for Research in the Atmosphere, Colorado State University, Ft. Collins, CO, USA

<sup>7</sup>Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

\*[Now at Jet Propulsion Laboratory, California Institute of Technology, CA, USA](#)

**Correspondence:** Brendan Byrne (brendan.k.byrne@jpl.nasa.gov)

**Abstract.** Interannual variations in temperature and precipitation impact the carbon balance of terrestrial ecosystems, leaving an imprint in atmospheric CO<sub>2</sub>. Quantifying the impact of climate anomalies on the net ecosystem exchange (NEE) of terrestrial ecosystems can provide a constraint to evaluate terrestrial biosphere models against, and may provide an emergent constraint on the response of terrestrial ecosystems to climate change. We investigate the spatial scales over which interannual variability in NEE can be constrained using atmospheric CO<sub>2</sub> observations from the Greenhouse Gases Observing Satellite (GOSAT). NEE anomalies are calculated by performing a series of inversion analyses using the GEOS-Chem model to assimilate GOSAT observations. Monthly NEE anomalies are compared to “proxies”, variables which are associated with anomalies in the terrestrial carbon cycle, and to upscaled NEE estimates from FLUXCOM. ~~Strong agreement is found in the timing of anomalies in the GOSAT flux inversions with~~ Statistically significant correlations ( $P < 0.05$ ) are obtained between posterior NEE anomalies and anomalies in soil temperature and FLUXCOM ~~. Strong correlations are obtained ( $P < 0.05$ ,  $R > R_{\text{NINO3.4}}$ ) in the tropics~~ NEE on continental and larger scales in the tropics, and in the northern extratropics on sub-continental scales during the summer ( $R^2 \geq 0.49$ ). ~~These results, in addition to a series of observing system simulation experiments that were conducted, provide evidence that GOSAT flux inversions can isolate anomalies in NEE on continental and larger scales. However, in both the tropics and northern extratropics, the agreement between the inversions and the proxies/FLUXCOM is sensitive to the flux inversion configuration. Our results suggest that regional scales are likely the minimum scales that can be resolved in the tropics using GOSAT observations, but obtaining robust NEE anomaly estimates on these scales may be difficult.~~ suggesting that GOSAT measurements provide a constraint on NEE IAV on these spatial scales. Furthermore, we show that GOSAT flux inversions are generally better correlated with the environmental proxies and FLUXCOM NEE IAV than NEE anomalies produced by a set of terrestrial biosphere models (TBMs), suggesting that GOSAT flux inversions could be used to evaluate TBM NEE fluxes.

## 1 Introduction

Organisms within terrestrial ecosystems have evolved to fit their climatic environment. Anomalous variations in temperature and precipitation about the mean climate can have significant impacts on the functioning of these organisms (Berry and Bjorkman, 1980; Gutschick and BassiriRad, 2003; Smith, 2011), which can be reflected in anomalies in the carbon balance of ecosystems. In fact, interannual variability (IAV) in the atmospheric growth rate of CO<sub>2</sub> is largely explained by changes in the carbon balance of terrestrial ecosystems in response to climate variability (Keeling et al., 1976a, b; Conway et al., 1994; Keeling et al., 1995; Battle et al., 2000). ~~A number of studies have taken advantage of this fact to estimate anomalies in net ecosystem exchange (NEE)~~ The largest driver of IAV in the carbon cycle is El Niño-Southern Oscillation (ENSO) variability, which most strongly impacts tropical ecosystems (Bacastow, 1976; Bacastow et al., 1980; Bousquet et al., 2000; Ciais et al., 2013). During the warm phase of ENSO (El Niño) large areas of tropical land become drier and warmer, leading to a net emission of CO<sub>2</sub> from the land to the atmosphere, which amplifies the atmospheric CO<sub>2</sub> growth rate. During the cold phase of ENSO (La Niña), much of the tropical land is cooler and wetter than average, leading to anomalously low CO<sub>2</sub> growth rates (Jones and Cox, 2005). Similarly, in extratropical regions, temperature and moisture anomalies drive variability in the carbon balance of ecosystems (Conway et al., 1994; Bousquet et al., 2000; Wunch et al., 2013; Zscheischler et al., 2014; He et al., 2018). Many studies have examined extreme heatwaves or droughts in the extratropics, such as the 2003 European heatwave (Ciais et al., 2005) and 2010 Russian heat wave and wildfires (Guerlet et al., 2013; Ishizawa et al., 2016). In these studies, ~~data assimilation methods are employed to estimate NEE anomalies consistent with measured variations in atmospheric cases, positive temperature anomalies and drought conditions result in an anomalous release of CO<sub>2</sub>~~ (Bousquet et al., 2000; Rödenbeck et al., 2003; Bruhwiler et al., 2003). ~~These studies have generally used from terrestrial ecosystems to the atmosphere.~~

Most previous studies that have investigated IAV in the carbon cycle using CO<sub>2</sub> measurements have employed measurements from the global network of in situ instruments for observational constraints (Bousquet et al., 2000; Rödenbeck et al., 2003; Bruhwiler et al., 2003). This network provides by far the longest direct record of atmospheric CO<sub>2</sub> measurements, with many sites functioning for decades. However, the spatial distribution of sites is inhomogeneous, with sites most densely located in North America and Europe and comparatively few elsewhere. Therefore, in situ observations from the global observation network are relatively insensitive to CO<sub>2</sub> fluxes over much of Asia and in the tropics (Byrne et al., 2017), where IAV is the largest. Recently, space-based observations of atmospheric CO<sub>2</sub> have provided expanded observational coverage for atmospheric CO<sub>2</sub>. One of the satellites, the Greenhouse Gases Observing Satellite (GOSAT), has been providing measurements of atmospheric CO<sub>2</sub> since 2009. With multiple years of measurements, it is now possible to investigate IAV in the carbon cycle with GOSAT data.

In this study, we investigate interannual ~~flux anomalies~~ anomalies in NEE estimated from GOSAT measurements using the ~~“flux inversion”~~ flux inversion method, wherein surface fluxes are estimated from atmospheric CO<sub>2</sub> measurements using a tracer transport model and Bayesian inverse methods. A series of flux inversions using the GEOS-Chem four-dimensional variational (4D-Var) data assimilation system (Henze et al., 2007) are performed with different spatial resolutions, prior fluxes

and prior error covariances. ~~We also examine the posterior fluxes from two~~ [NEE anomalies produced by the GEOS-Chem flux inversions are contrasted with two independent](#) publicly available flux inversion estimates, the GOSAT Level 4 product (Maksyutov et al., 2013) and CarbonTracker, version CT2016 (Peters et al., 2007, with updates documented at <http://carbontracker.noaa.gov>) which is a flux inversion that assimilates CO<sub>2</sub> observations from the surface network. ~~Posterior anomalies~~ [In addition, we compare the flux inversion NEE anomalies with NEE anomalies simulated by several terrestrial biosphere models \(TBMs\). One of the primary goals of atmospheric flux inversions is to provide better constraints on NEE to evaluate these models. Therefore, it is necessary to first determine the ability of TBMs to represent IAV.](#)

[Anomalies](#) in NEE from the inversions [and TBMs](#) are compared with “proxies”, variables that are known to be closely associated with IAV in the carbon cycle. Agreement between the anomalies in ~~the inversions~~ [NEE](#) and proxies provides corroborating evidence that the inversions/[TBMs](#) are correctly recovering anomalies in NEE (Deng et al., 2016; Mabuchi et al., 2016; Liu et al., 2017). Three proxies are examined: soil temperature ( $T_{soil}$ ) anomalies from the [Modern-Era Retrospective analysis for Research and Applications \(MERRA-2\)](#) reanalysis (Reichle et al., 2011, 2017), the Monthly Self-calibrated Palmer Drought Severity Index (scPDSI) (Dai, 2017), and solar-induced chlorophyll fluorescence (SIF) observed by GOME-2 (Joiner et al., 2016). We also use flux data from FLUXCOM ([www.fluxcom.org](http://www.fluxcom.org)), which provides data-driven NEE anomaly estimates (Tramontana et al., 2016; Jung et al., 2017).

~~Anomalies in temperature and water availability are closely linked to anomalies in terrestrial ecosystems. On the local scales of FLUXNET sites (Baldocehi et al., 2001), temperature and precipitation have both been shown to be major controls on NEE (see Baldocehi et al. (2018) for a review). On regional and larger scales, stronger correlations have been found with temperature anomalies than with precipitation anomalies (Wang et al., 2013; Jung et al., 2017), particularly in the tropics. Jung et al. (2017) suggest that this is partially due to sub-regional scale spatial variability in water availability anomalies that compensate, thereby reducing the influence of these anomalies on larger scales, while temperature anomalies are generally more spatially coherent. Nevertheless, both temperature and water availability anomalies strongly influence NEE anomalies over a wide range of scales. The largest driver of IAV in the carbon cycle is El Niño-Southern Oscillation (ENSO) variability, which most strongly impacts tropical ecosystems (Bacastow, 1976; Bacastow et al., 1980; Bousquet et al., 2000; Ciais et al., 2013). During the warm phase of ENSO (El Niño) large areas of tropical land become dryer and warmer, leading to a net emission of from the land to the atmosphere, which amplifies the atmospheric growth rate. During the cold phase of ENSO (La Niña), much of the tropical land is cooler and wetter than average, leading to anomalously low growth rates (Jones and Cox, 2005). In the extratropics, there is also significant variability in the carbon balance of terrestrial ecosystems related to temperature and moisture anomalies (Conway et al., 1994; Bousquet et al., 2000). Wunch et al. (2013) show that the summer minima in column-averaged dry-air mole fraction of ( $X_{CO_2}$ ) observed at northern midlatitude Total Carbon Column Observing Network (TCCON) sites is correlated with surface temperature, indicating that midlatitude positive temperature anomalies correspond to reduced uptake by the northern extratropical biosphere during the growing season. Similarly, He et al. (2018) show that GOSAT  $X_{CO_2}$  anomalies are correlated with indicators surface environmental parameters (such as temperature and drought index). Many studies have examined extreme heatwaves or droughts in the extratropics, such as the 2003 European heatwave (Ciais et al., 2005) and 2010 Russian heat wave and wildfires (Guerlet et al., 2013; Ishizawa et al., 2016). In these cases, positive~~

temperature anomalies and drought conditions result in a release of from terrestrial ecosystems to the atmosphere. Zscheischler et al. (2014) that relatively few extreme events dominate anomalies in gross primary productivity (GPP), and likely NEE. Due to the large seasonal cycle of temperature, precipitation and insolation in the extratropics, the relationship between anomalies in NEE and the proxies is likely a function of time of year. We focus our study of the northern extratropics to the Northern Hemisphere summer (JJA).

SIF is the emission of radiation by chlorophyll during photosynthesis and thus provides a measure of GPP (Papageorgiou and Govindjee, 2008). Therefore, reduced GPP is associated with reduced SIF, and vice-versa. The relationship to anomalies in NEE is less direct because GPP and ecosystem respiration anomalies are highly correlated (Baldochi et al., 2018). Therefore, the extent to which SIF anomalies and NEE anomalies should be correlated is not well understood. One study, Shiga et al. (2018b), shows that SIF can be used to inform the spatiotemporal distribution of NEE over North America.

Upscaled NEE estimates from eddy covariance measurements at flux towers can be used to generate an observation-based estimate of NEE anomalies at regional to continental scales. Kondo et al. (2015) compared the GOSAT L4 product and empirical eddy flux upscaling and found similar responses to climate anomalies in temperate and boreal regions, while poorer agreement was found in the tropics. Here, we use upscaled NEE estimates from FLUXCOM that are generated using upscaling approaches based on machine learning methods that integrate FLUXNET site level observations of fluxes, satellite remote sensing, and meteorological data (Framontana et al., 2016; Jung et al., 2017). For this study, upscaled fluxes generated using multivariate regression splines (MARS) are used. Similar results were found for other upscaling algorithms.

It is important to acknowledge that none of these proxies (or FLUXCOM) should be expected to be perfectly correlated with the true NEE anomalies. Therefore, when there is disagreement between the inversions and proxies, it unclear whether this should be attributed to the inversion NEE or the proxy. Comparisons of flux inversions with the proxies are most useful for identifying “positive” results for which the assimilation of atmospheric observations has introduced a strong correlation with the proxies. However, these comparisons are less useful for identifying the limits of the inversions with “negative” results, in which the null hypothesis (no correlation) cannot be rejected.

In addition to comparing our flux inversions with the proxies and the FLUXCOM data, we also compare several terrestrial biosphere models (TBMs) with the proxies and the FLUXCOM data. TBMs simulate GPP and ecosystem respiration, and therefore provide estimates of NEE. TBMs are widely applied to simulate projections of the future carbon cycle, however, different models show large disagreements on the relative importance of different processes driving the uptake (Huntzinger et al., 2017). One of the primary goals of atmospheric flux inversions is to provide better constraints on NEE to evaluate these models. Therefore, it is useful to determine whether the agreement between flux inversions and the proxies is closer than the agreement between TBMs and the proxies.

This paper has three study has two main objectives. The first is to quantify the agreement correlation between GOSAT flux inversions and the flux proxies. This will be useful for identifying the utility of using proxies to corroborate flux inversions results. The second is to, and determine the spatial scales over which the GOSAT inversions constrain flux anomalies. GOSAT observations are expected to best constrain fluxes on large scales, such as the entire tropics. As scales decrease, finer scale structures in the atmospheric fields are required to constrain fluxes, the smallest scales at which GOSAT observations provide

useful constraints on NEE anomalies is currently unclear. We quantify the ability of GOSAT flux inversions to quantify NEE anomalies over a range of spatial scales by, first, examining the agreement between the inversions and proxies over a range of spatial scales and, second, examine the ability of GOSAT inversions to recover true flux anomalies by performing a series of Observing System Simulation Experiments (OSSEs). Monthly anomalies in the tropics are examined throughout the year while anomalies in the northern extratropics are examined during the summer (June-July-August, JJA). The third monthly NEE anomalies in the tropics and summer anomalies in the northern extratropics. This is performed by examining the spatial scales over which there are correlations between the NEE anomalies and proxies/FLUXCOM anomalies. The second objective is to quantify the sensitivity of the results for the first two objectives inversion – proxy correlations to the inversion setup. This is investigated with a series of GOSAT flux inversions with different model resolution, prior fluxes, and prior error covariances.

This paper is structured as follows. In Sect. 2, we describe the datasets used, flux inversions performed, and how anomalies are calculated. Sect. 3 presents the results of our analysis. Flux inversion NEE anomalies are compared with the proxies in the tropics and northern extratropics individually. We then present an OSSE a series of Observing System Simulation Experiments (OSSEs) to examine the smallest spatial scales for which NEE anomalies can be recovered from GOSAT observations. Sect. 4 discusses the agreement in anomalies between the GOSAT flux inversions and proxies, the scales constrained by GOSAT flux inversions, and the sensitivity of these results to the inversion setup. Finally, conclusions are given in Sect. 5.

## 2 Data and methods

### 2.1 FLUXCOM NEE Data

FLUXCOM remote sensing and meteorological data (RS+METEO) products are generated at  $0.5^\circ \times 0.5^\circ$  spatial resolution using upscaling approaches based on machine learning methods that integrate FLUXNET (<https://fluxnet.ornl.gov/>, ?) site level observations, satellite remote sensing, and meteorological data (Tramontana et al., 2016; Jung et al., 2017). Explanatory variables from remote Remote sensing measurements are averaged used to produce a mean seasonal cycle (Tramontana et al., 2016), such that all NEE IAV is introduced by and the NEE IAV signal in this product results from the driving reanalysis (NCEP CRU). In particular, NEE IAV is driven by air temperature, incoming global radiation combined with the mean seasonal cycle of NDVI, and model-based water availability index. Jung et al. (2017) generate NEE products using several machine learning methods. We downloaded these products from the Data Portal of the Max Planck Institute for Biochemistry (<https://www.bgc-jena.mpg.de>). We find that the different algorithms generally give similar results, therefore we only present results using the multivariate regression spline (MARS) NEE in this study.

### 2.2 Proxies

#### 2.2.1 Dai Global Palmer Drought Severity Index

The monthly self-calibrated Palmer Drought Severity Index (scPDSI) (Dai, 2017) provides a measure of drought severity on a  $2.5^\circ \times 2.5^\circ$  grid. The scPDSI is computed using observed monthly surface air temperature and precipitation and provides a

measure of surface aridity anomalies and changes on seasonal to longer time scales (Dai et al., 2004; Dai, 2011). We note that scPDSI may not be a good proxy of soil moisture content over the high latitudes ( $>50^\circ$ ).

### 2.2.2 SIF

We use the monthly gridded “SIF daily average” product from the NASA Level 3 GOME-2 version 27 (V27) terrestrial chlorophyll fluorescence data (NASA-SIF, 2016; Joiner et al., 2013, 2016). SIF anomalies are multiplied by negative one to change the sign of the anomalies, so that positive correlations will be obtained if negative SIF anomalies are correlated with positive NEE anomalies (emission of  $\text{CO}_2$  to the atmosphere).

### 2.2.3 Soil temperature

For the soil temperature proxy, we use soil temperatures from the MERRA-2 (Reichle et al., 2011, 2017) reanalysis. ~~Specifically~~Specifically, we use the average soil temperature over levels 1–3 (TSOIL1, TSOIL2, and TSOIL3), which reaches a depth of 0.73 m.

### 2.2.4 NINO 3.4 index

For the phase of ENSO, we use the sea surface temperature (SST) anomaly in the NINO 3.4 region ( $5^\circ\text{S}$ – $5^\circ\text{N}$ ,  $120^\circ\text{S}$ – $170^\circ\text{N}$ ) of the tropical Pacific Ocean. This region has been widely used to diagnose ENSO activity. The SST values are calculated from the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST) dataset. The SST anomalies were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) website (<https://www.esrl.noaa.gov>).

## 2.3 Inversion analyses

### 2.3.1 CarbonTracker

We use optimized NEE from the NOAA’s CarbonTracker, version CT2016 (Peters et al., 2007, with updates documented at <http://carbontracker.noaa.gov>). CT2016 optimizes NEE by assimilating in situ observations of boundary layer atmospheric  $\text{CO}_2$ . It employs the ensemble Kalman filter approach to assimilate  $\text{CO}_2$  with atmospheric chemical transport simulated by the TM5 offline atmospheric model (Krol et al., 2005). For CT2016, TM5 is driven by ERA-Interim assimilated meteorology from the European Centre for Medium-Range Weather Forecasts (ECMWF), with a horizontal resolution of  $3^\circ \times 2^\circ$  globally and  $1^\circ \times 1^\circ$  in a nested grid over North America. CT2016 also has IAV in biomass burning. Therefore, when analyzing posterior IAV in CT2016 we examine the IAV in NEE alone (referred to as CT2016) and IAV due to NEE and biomass burning combined (referred to as CT2016<sub>w/BB</sub>).

### 2.3.2 GOSAT level 4 data

We use the GOSAT level 4 data product (Maksyutov et al., 2013) produced by the National Institute for Environmental Studies (NIES). This product is produced by assimilating NIES Level 2 retrievals of  $\text{XCO}_2$  into the NIES global atmospheric tracer

transport model (NIES-TM) to optimize monthly CO<sub>2</sub> fluxes for 64 sub-continental regions. The Vegetation Integrative Simulator for Trace gases (VISIT), a prognostic biosphere model (Ito, 2010; Saito et al., 2014), is used to generate prior biospheric fluxes for the inversion analyses. The GOSAT L4 product also has IAV in biomass burning. Therefore, when analyzing posterior IAV, we examine IAV in NEE alone (referred to as GOSAT L4) and IAV due to NEE and biomass burning combined (referred to as GOSAT L4<sub>w/BB</sub>).

### 2.3.3 GEOS-Chem

We perform a series of flux inversions using the GEOS-Chem 4D-Var assimilation system (Henze et al., 2007). The GEOS-Chem forward model ([www.geos-chem.org](http://www.geos-chem.org)) is a global 3-D chemical transport model driven by assimilated meteorology from the Goddard Earth Observing System (GEOS-5) of the NASA Global Modeling and Assimilation Office (GMAO). The native resolution of the model is  $0.5^\circ \times 0.67^\circ$  with 72 vertical levels from the surface to 0.01 hPa, but we run the model at lower resolution (either  $2^\circ \times 2.5^\circ$  or  $4^\circ \times 5^\circ$ , depending on the inversion) with 47 vertical layers. Our model configuration is based on the configuration of Nassar et al. (2011). To optimize surface fluxes, the 4D-Var cost function is minimized as described in Deng et al. (2014) to retrieve monthly scaling factors for prior ocean and terrestrial biosphere fluxes in each grid cell. We use an assimilation window of nine months and keep posterior fluxes from the first six months, then shift the inversion window forward six months. Using this method, we optimize NEE spanning 2010–2013 (in addition to a six month spin up inversion starting in July 2009). Monthly ocean fluxes are from Takahashi et al. (2009), anthropogenic emissions are from Andres et al. (2016), and biomass burning emissions are from the Global Fire Emission Database GFEDv3 (van der Werf et al., 2006). We repeat the 2010 GFEDv3 biomass burning emissions for all years so that there is no prior NEE IAV. Error covariance matrices are taken to be diagonal, such that there are no spatial or temporal covariances. Prior errors are assigned as a percentage of the prior flux estimate rather than an absolute value. We assign 16% error to fossil fuels and 38% error to biomass burning following Deng et al. (2014).

The GEOS-Chem flux inversions performed in this study are shown in Table 1. The flux inversions are performed with different model configurations to examine the sensitivity of the results to the inversion setup. We perform inversions at two spatial resolutions,  $2^\circ \times 2.5^\circ$  and  $4^\circ \times 5^\circ$ . The spatial resolution is varied to examine whether changes in model transport significantly impact our results. It has previously been shown that there are significant differences in tracer transport as model resolution is decreased in GEOS-Chem (Yu et al., 2018; Stanevich, 2018). In particular, Stanevich (2018) show that resolution-induced biases of up to 30% can be introduced on the scale of TransCom regions for  $4^\circ \times 5^\circ$  relative to  $2^\circ \times 2.5^\circ$  for atmospheric methane (CH<sub>4</sub>) inversions.

The prior error statistics are varied between inversions. The prior error covariance provides a metric of the uncertainty in the prior fluxes. If prior fluxes are well known then small errors are applied. If they are poorly known then large prior errors are applied and the observations will have a larger impact on the posterior fluxes. However, in general, atmospheric CO<sub>2</sub> observations underconstrain the fluxes and additional regularization considerations are required. To prevent overfitting of assimilated observations, prior flux errors are typically tighter than the true uncertainty in NEE fluxes. Therefore, the motivation for varying prior errors in this study is to examine the sensitivity of the posterior flux-NEE anomalies to these prior constraints.

Finally, the prior NEE fluxes are varied between flux inversions. For all GEOS-Chem inversions, the prior NEE fluxes are based on the posterior fluxes from CT2016. CT2016 fluxes are used because they are informed by atmospheric CO<sub>2</sub> observations, and thus provide a seasonal cycle of NEE which is closer to observed atmospheric CO<sub>2</sub> than a TBM forward run (Byrne et al., 2018). Using prior fluxes which are closer to the observed atmospheric CO<sub>2</sub> then justifies tighter prior flux error covariances. We use two different setups of the CT2016 posterior fluxes in the inversions. For four inversions we remove the NEE IAV from the CT2016 fluxes. To do this, the fluxes are averaged over the four years (2010–2013) to generate a mean seasonal cycle. We then repeat this climatology of NEE fluxes for each year of the inversion. The reason for removing prior NEE IAV is so that all posterior NEE anomalies will be introduced through the assimilation of GOSAT observations. This setup is different from many previous flux inversion studies which have included NEE IAV in the prior fluxes. Therefore, to examine the sensitivity of the posterior NEE IAV to prior NEE IAV, we also perform two inversions that employ three-hourly CT2016 NEE fluxes over 2010–2013 unchanged from those available at <http://carbontracker.noaa.gov>, other than spatial interpolation to fit our grid, so that NEE IAV is present on the prior NEE for these inversions. The inversions are given names with a subscript following the convention “model resolution – percentage error applied to prior fluxes – presence of prior NEE IAV”, such that, an inversion analysis at  $4^\circ \times 5^\circ$  resolution with 100% uncertainty applied to prior fluxes and with prior NEE IAV is named “GC<sub>4×5–100%–IAV</sub>.”

For GOSAT observations, we use version v3.5 of the NASA Atmospheric CO<sub>2</sub> Observations from Space (ACOS) GOSAT lite files from the CO<sub>2</sub> Virtual Science Data Environment (<https://co2.jpl.nasa.gov/#mission=ACOS>). Information on the ACOS retrieval algorithm is available in O’Dell et al. (2012) and Crisp et al. (2012). We selected all measurements from the TANSO-FTS shortwave infrared (SWIR) channel, including ocean glint, high gain and medium gain nadir, which pass the quality flag requirement and have warn levels less than or equal to 15. We generate “super-obs” from the GOSAT retrievals by aggregating the observations to the grid size of our inversion. We generate error estimates using the method described by Kulawik et al. (2016). The reduction in error with aggregation can be calculated using the expression  $error^2 = a^2 + b^2/n$ , where  $a$  represents systematic errors that do not decrease with averaging,  $b$  represents random errors that decrease with averaging, and  $n$  represents the number of satellite observations that are averaged (Kulawik et al., 2016). Kulawik et al. (2016) give  $a = 0.8$  ppm and  $b = 1.6$  ppm as mean Northern Hemisphere geometric (co-located) values for GOSAT, and these are the values that we use.

### 2.3.4 Observing system simulation experiments

Five OSSEs are performed, for which pseudo-data are generated by simulating atmospheric CO<sub>2</sub> with GEOS-Chem at  $4^\circ \times 5^\circ$  spatial resolution and with year-specific NEE from the Joint UK Land Environment Simulator (JULES). The GEOS-Chem CO<sub>2</sub> distribution is sampled according to the GOSAT observational coverage. We generate pseudo XCO<sub>2</sub> using the GOSAT averaging kernel weighting and apply random errors to the XCO<sub>2</sub> pseudo-obs consistent with the error estimates described in Sect. 2.3.3. The inversion configuration for three of the OSSEs is identical to GC<sub>4×5–44%</sub>, GC<sub>4×5–100%</sub>, and GC<sub>4×5–100%–IAV</sub>, which use the posterior CT2016 fluxes as their prior NEE (see Table 1). These OSSEs are referred to as OSSE<sub>CT2016–44%</sub>, OSSE<sub>CT2016–100%</sub>, and OSSE<sub>CT2016–100%–IAV</sub>, respectively. Two more OSSEs use the same setup as GC<sub>4×5–44%</sub> and



$GC_{4 \times 5-100\%}$ , except that for these we use the 2010-2013 mean NEE fluxes from JULES as the prior fluxes. These two OSSEs are referred to as  $OSSE_{JULES-44\%}$  and  $OSSE_{JULES-100\%}$ .

## 2.4 Terrestrial biosphere models

### 2.4.1 JULES

5 JULES is a community land surface model that has evolved from the UK Met Office Surface Exchange Scheme. Phenology in JULES affects leaf growth rates and timing of leaf growth/senescence based on temperature alone (Clark et al., 2011). Vegetation cover is predicted based on nine plant functional types that compete for space based on their relative productivity and height but are excluded from growing on agricultural land, based on a fraction of agriculture in each grid cell (Harper et al., 2018). CRU-NCEP was used as model forcing data.

### 10 2.4.2 VISIT

VISIT is a prognostic biosphere model (Ito, 2010; Saito et al., 2014) that simulates carbon exchanges between the atmosphere and biosphere and among the carbon pools within terrestrial ecosystems at a daily time step. Modeling of plant  $CO_2$  assimilation in VISIT is based on a model of light extinction in the canopy, following the formulation of Monsi and Saeki (1953). Autotrophic respiration is formulated as the sum of growth respiration and maintenance respiration. Growth respiration is simulated as the cost to produce new biomass, while maintenance respiration is represented as a function of ground surface temperature. Heterotrophic respiration is the sum of respiration from two layers, litter and humus, which is regulated by soil temperature and soil moisture at each depth. VISIT is driven by reanalysis/assimilation climate datasets provided by the Japan Meteorological Agency (JMA): the Japan 25-year reanalysis (JRA-25)/JMA Climate Data Assimilation System JCDAS (Onogi et al., 2007) for the period 1979–present.

### 20 2.4.3 Carnegie-Ames-Stanford Approach (CASA) Global Fire Emissions Database (GFED) Carbon Monitoring System (CMS) model

The version of the CASA model used here, referred to as CASA CMS, was modified from Potter et al. (1993) as described in Randerson et al. (1996), van der Werf et al. (2006) and Liu et al. (2014). It is driven by MERRA reanalysis and satellite-observed Normalized Difference Vegetation Index (NDVI) to track plant phenology. These flux estimates were computed at monthly time steps with  $0.5^\circ$  horizontal resolution. Monthly NEE fluxes were downloaded from the CarbonTracker ftp (<ftp://aftp.cmdl.noaa.gov/products/carbontracker/co2/CT2016/fluxes/priors/>).

### 2.4.4 CASA GFED 4.1

The version of the CASA model used here, CASA GFED 4.1, was modified from Potter et al. (1993) as described in van der Werf et al. (2017). It is driven by ECMWF reanalysis and satellite-observed NDVI to track plant phenology. These flux esti-

mates were computed at monthly time steps with  $0.25^\circ$  horizontal resolution. Monthly NEE fluxes were downloaded from the CarbonTracker ftp (<ftp://aftp.cmdl.noaa.gov/products/carbontracker/co2/CT2016/fluxes/priors/>).

## 2.5 Anomalies and correlations

Monthly anomalies are calculated by subtracting the mean 2010-2013 value for a given month from the monthly value for a specific year. For example, the NEE anomaly for a given month and year is calculated using:

$$ANOM[year, month] = NEE[year, month] - \frac{1}{4} \sum_{i=2010}^{2013} NEE[i, month]. \quad (1)$$

Anomalies are calculated over a range of spatial scales. In each case, the quantity of interest is first averaged into a spatial mean for each month, then anomalies are calculated. The same procedure is followed for JJA anomalies except that the anomaly is calculated over the entire three month period instead of for a single month. [It is worth noting that four years is a relatively short period to define a climatology, and some modes of climate variability occur on longer timescales. Ideally, a longer time period would be used to calculate a climatology, but we are limited by the availability of GOSAT data in this study.](#)

In the tropics, temporal correlations are performed to quantify agreement between NEE anomalies and proxy/FLUXCOM anomalies. We want to test the hypothesis that the assimilation of  $\text{CO}_2$  observations will significantly increase the correlation between the posterior NEE IAV and the proxies relative to the prior NEE IAV and the proxies. We choose a null hypothesis in which the correlation is zero. This is the correct null hypothesis for flux inversions for which the prior NEE fluxes have no IAV. In flux inversions for which there is IAV in the prior NEE, the correlation between the proxies and prior NEE IAV should be used as the null hypothesis. However, this would be a significantly more difficult null hypothesis to test, so for simplicity we choose a null hypothesis of zero correlation for all cases. This is equivalent to testing whether the posterior NEE IAV is significantly correlated with the proxies, regardless of the prior IAV. The threshold for rejection of the null hypothesis ( $\alpha$ ) is chosen to be 0.05, such that the null hypothesis is rejected if the P-value (P) is less than 0.05. We acknowledge that this  $\alpha$  threshold is largely arbitrary but is widely used in the literature (Benjamin et al., 2018; Lakens et al., 2018). Throughout the manuscript, correlations that satisfy this criterion are called “strong”. In most cases a second test is performed, in which we test if the correlation between the flux inversion NEE IAV and the proxy is greater than that between the NINO 3.4 index and the proxy, and conclude that the inversion and proxy only show good agreement if both of these thresholds are met. The coefficient of correlation is referred to as  $R$ .

We also perform a series of linear regressions. In the tropics, linear regressions are performed after aggregating over all tropical land, such that the regression is performed on a single 48 point time series. In the northern extratropics, linear regressions are performed for the set of four JJA anomalies across five sub-continental regions resulting in a 20 point dataset. For all regressions the y-intercept is close to zero, and thus is not reported. The slope of the regressions and coefficient of determination ( $R^2$ ) are reported.

### 3 Results

#### 3.1 Tropics

~~Monthly~~ This section presents the correlations between monthly NEE anomalies and proxy/FLUXCOM anomalies in the tropics ~~are examined over a range of spatial scales. The anomalies are aggregated to~~. The analysis is performed at five different ~~spatial scales:  $4^\circ \times 5^\circ$ ,  $8^\circ \times 10^\circ$ , sub-continental regions (shown in Fig. 1), continents, and the entire tropics. Figure 2 shows the mean correlation coefficient between the inversions /FBMs and proxies/FLUXCOM on scales ranging from  $4^\circ \times 5^\circ$  grid cells to the entire tropics. Correlations between the NINO 3.4 index and flux proxies are also shown over the range of spatial scales. It is important to consider correlations between the inversions/TBMs and proxies/FLUXCOM with the influence of ENSO variability in mind, as ENSO is the primary driver of large scale NEE IAV these scales in the tropics. Therefore, to understand how well the flux inversions are capturing NEE IAV it is useful to contrast correlations between the inversion and proxies to correlations between the NINO 3.4 index and proxies.~~

The correlation between posterior NEE anomalies and proxy/FLUXCOM anomalies increase with spatial aggregation (Fig. 2). This is expected as atmospheric CO<sub>2</sub> observations are expected to best constrain fluxes on large spatial scales, such as the entire tropics. ~~As scales decrease, the~~ At smaller spatial scales, the atmospheric signal from variations in the fluxes become weaker and more difficult to constrain with the atmospheric CO<sub>2</sub> observations. ~~Correlations between the proxies and the NINO 3.4 index also increase with aggregation. This is because the NINO 3.4 index reflects the large scale ENSO-driven variability in the tropics. Therefore, increasing correlation with aggregation for the NINO 3.4 index is a reflection of the large-scale variability having a larger impact.~~

~~To categorize the agreement between the flux inversions and the proxies/FLUXCOM, we state that the flux inversions agrees with a proxy on a given scale only if the correlation is strong ( $P < 0.05$ ) and greater than the correlation of the proxy/FLUXCOM with~~ In general, the GEOS-Chem GOSAT inversions show strong correlations ( $P < 0.05$ ) with FLUXCOM and the NINO 3.4 index ( $R > R_{\text{NINO3.4}}$ ). For the GEOS-Chem inversions, this occurs at regional on sub-continental regions and larger scales ~~for correlations with FLUXCOM NEE and at continental and larger scales for  $T_{\text{soil}}$ . For the GOSAT L4 inversion, the correlation only reaches this threshold for~~ and with  $T_{\text{soil}}$  at the largest aggregation scale. These results suggest ~~that GOSAT observations provide flux information,  $\text{scPDSI}$  on continental and larger scales, while regional-scale constraints may be possible. The fact that the~~ We do not find strong correlations with GOME-2 SIF on any scale in the tropics.

The correlation coefficient is variable between GOSAT inversions ~~indicates~~, suggesting that the agreement between posterior fluxes and the proxies/FLUXCOM is sensitive to the inversion configuration.

~~We investigate the influence of the inversion configuration by comparing the correlations for the six GEOS-Chem inversion.~~ The  $2^\circ \times 2.5^\circ$  inversions generally show slightly better agreement with the proxies/FLUXCOM than the  $4^\circ \times 5^\circ$  inversions ~~at regional~~ for sub-continental regions and continental scales. ~~This could be due to improved transport with higher spatial resolution, however, other aspects of the inversion were changed such as the aggregation of assimilated observations and prior error covariances, which may have also influenced the results. The influence of prior NEE IAV can be evaluated by comparing the  $4^\circ \times 5^\circ$  inversions with and without prior NEE IAV. Correlations are stronger for the inversions without NEE IAV at regional~~

for sub-continental regions and continental scales. This suggests that the presence prior NEE IAV can degrade posterior NEE IAV and is discussed in more detail in Sec. 4.4.3. The influence of prior error covariances can be evaluated by comparing the inversions with small (44% for  $4^\circ \times 5^\circ$  and 66% for  $2^\circ \times 2.5^\circ$ ) and large (100% for  $4^\circ \times 5^\circ$  and 200% for  $2^\circ \times 2.5^\circ$ ) prior error. Larger prior errors generally result in larger correlations on regional for sub-continental regions and larger scales. Large prior errors means that more movement away from the prior during the inversion is allowed, therefore, better agreement with larger prior errors suggests that the GOSAT data information content is sufficiently large that loose prior errors can be applied without degrading the posterior results by over-fitting the observations.

The GOSAT L4 product tends to give weaker correlations with the proxies than the GEOS-Chem flux inversions, which is likely due to differences in the inversion configuration. For CT2016, strong correlations that are greater than those for the NINO 3.4 index are only obtained for SIF and only on the scale of the entire tropics. The reason for the strong correlation with SIF is unclear, but it could be a result of the fact that NEE IAV in one of the CT2016 priors (CASA GFED 4.1) is strongly correlated with SIF. The poorer agreement between CT2016 and which assimilates surface in situ and flask measurements, correlations with the proxies/FLUXCOM than for GOSAT inversions suggests that the network of surface observations does not provide sufficient information to constrain tropical fluxes. However, it is also possible that the inversion setup could play a role. are generally weaker than for the GOSAT flux inversions.

For the TBMs, correlations are highly model dependent. Of the models, JULES shows the best agreement with the proxies/FLUXCOM. JULES shows strong correlations greater than for the NINO 3.4 index ( $P < 0.05$ ) at all scales for FLUXCOM NEE, regionally and over for sub-continental regions and the entire tropics for  $T_{soil}$ , and regionally sub-continental regions for scPDSI. These results suggest that JULES predicts NEE anomalies in the tropics as well as the GOSAT inversion on continental and larger scales, and may be better at regional and smaller scales. This suggests that it may be challenging to use GOSAT flux inversions to evaluate IAV in JULES NEE. For the other models, less agreement is seen with the proxies/FLUXCOM. The one exception is CASA GFED 4.1 which shows strong correlations with SIF at all scales. This may be due to the fact that this model assimilates greenness indices to estimate GPP fluxes. Anomalies in the greenness indices are likely well correlated with SIF anomalies, therefore, if anomalies in CASA NEE are driven by anomalies in GPP, it may explain the strong correlation.

We now investigate the magnitude of tropical NEE IAV in the inversions and the TBMs. The magnitude of NEE IAV relative to the proxies/FLUXCOM can be obtained by performing linear regressions of the inversion/TBM NEE anomalies against proxy/FLUXCOM anomalies. Linear regressions are only calculated for the scale of the entire tropics, where the inversions and proxies/FLUXCOM agreed best. Table 2 shows the slope and coefficient of determination ( $R^2$ ) for the regressions. There is a large amount of variability in the slopes between inversions/TBMs for each proxy/FLUXCOM. The GOSAT inversions inversion regressions are quite consistent with each other relative to those for CT2016 and the TBMs. The GOSAT inversions give slopes of 1.03–2.10 for FLUXCOM and 0.061–0.12 for  $T_{soil}$ . Comparing the (note that FLUXCOM NEE is known to underestimate t The largest differences between GEOS-Chem inversions, the largest differences in the slopes inversions are related to the magnitude of the prior error covariances. Looser prior constraints result in slopes that are 30–80% larger. There are also large differences in the magnitude of posterior NEE IAV between the inversions with and without prior NEE IAV. For example, the slopes for the regression between FLUXCOM and the  $4^\circ \times 5^\circ$  GEOS-Chem inversions with prior anomalies are 25–50% larger

than for GEOS-Chem inversions without prior NEE IAV. The GOSAT L4 product gives slopes which are consistent with the GEOS-Chem inversions. Furthermore, the agreement between the GOSAT L4 product and proxies (or FLUXCOM) is not sensitive to the inclusion of biomass burning. For CT2016, the best agreement is found with  $T_{soil}$  ( $0.24 \leq R^2 \leq 0.27$ ), for which CT2016 gives a smaller slope than the GOSAT inversions. The agreement between CT2016 and proxies/FLUXCOM is not sensitive to the inclusion of biomass burning. For the TBMs, JULES gives good fits with  $T_{soil}$  ( $R^2 = 0.56$ ) and FLUXCOM ( $R^2 = 0.47$ ) and gives slopes that are similar in magnitude to the flux inversions. The rest of the TBMs have  $R^2$  that are too small to make meaningful comparisons. [From this analysis we cannot make conclusions about the accuracy of the magnitude of IAV, instead this is addressed with a series of OSSEs in Sec. 3.3.](#)

### 3.1.1 Detailed analysis of $GC_{2 \times 2.5-200\%}$

~~We examine the agreement between the  $GC_{2 \times 2.5-200\%}$  inversion and the proxies/FLUXCOM in the tropics in more detail. Figure ?? (left column) shows the correlation coefficient for each grid cell between the  $GC_{2 \times 2.5-200\%}$  NEE anomalies and the proxy/FLUXCOM anomalies. There are broad positive correlations with the NINO 3.4 index across Central and South America, tropical and southern Africa, and much of the Asia-Pacific. Generally, positive correlations are present between  $GC_{2 \times 2.5-200\%}$  and SIF, sePDSI,  $T_{soil}$ , and FLUXCOM NEE in the Americas, southern Africa, and the Asia-Pacific. Figure ?? (center column) shows the correlation coefficient between the NINO 3.4 index and the proxies over the tropics. Generally, the proxies show strong correlations with the NINO 3.4 index in many of the same regions for which these proxies show strong correlations with  $GC_{2 \times 2.5-200\%}$ . This suggests that grid-scale correlations between  $GC_{2 \times 2.5-200\%}$  and the proxies may be a reflection of the large-scale anomalies across the tropics and do not necessarily imply that the inversion is able to isolate the spatial footprint of ENSO-driven flux anomalies on smaller scales. Alternatively, it is also possible that the proxies themselves do not correlate well with the true NEE at these scales.~~

~~We examine whether  $GC_{2 \times 2.5-200\%}$  is able to isolate flux anomalies that are separate from the large-scale tropical signal by comparing NEE anomalies for FLUXCOM NEE and  $GC_{2 \times 2.5-200\%}$  as a function of time. First, we aggregate  $GC_{2 \times 2.5-200\%}$  IAV and FLUXCOM NEE anomalies to the entire tropics and the following continental-scale regions: the Americas, Africa plus the Middle East, and the Asia-Pacific plus the Indian sub-continent (Fig. 1). Figure ?? shows  $GC_{2 \times 2.5-200\%}$  and FLUXCOM NEE anomalies as a function of time over the entire tropics and the continental-scale regions. We show raw and smoothed (3-month running mean) monthly NEE anomalies as a function of time. Over the entire tropics, FLUXCOM and  $GC_{2 \times 2.5-200\%}$  are highly correlated ( $R^2 = 0.69$ ) (which is shown in Fig. 2). On continental scales, the agreement between FLUXCOM and  $GC_{2 \times 2.5-200\%}$  is variable, ranging from  $R^2 = 0.08$  for Africa plus the Middle East to  $R^2 = 0.61$  for the Americas. All correlations improve after smoothing, suggesting that monthly scale variations are not correctly represented in  $GC_{2 \times 2.5-200\%}$ , FLUXCOM NEE, or both. We attempt to isolate anomalies specific to each continent by removing the large-scale anomaly across the entire tropics. This is done by subtracting a mean tropical anomaly (scaled to have the same variance as the continental anomaly) from the continental anomaly using the following equation:-~~

$$\text{DIFF}_{\text{continent-tropics}} = \text{ANOM}_{\text{continent}} - \text{ANOM}_{\text{tropics}} \times \frac{\text{STD}(\text{ANOM}_{\text{continent}})}{\text{STD}(\text{ANOM}_{\text{tropics}})},$$

where  $STD()$  represents standard deviation.  $DIFF_{\text{continent-tropics}}$  provides an estimate of anomalies in NEE for a given continent that are not associated with the large-scale ENSO-driven anomalies across the tropics.  $DIFF_{\text{continent-tropics}}$  is shown for each continent in Fig. ??e,h,k. The magnitude of the anomalies are reduced after removing the tropical mean anomalies. Positive correlations are obtained for the Americas ( $R^2 = 0.18$ ), Africa plus the Middle East ( $R^2 = 0.07$ ), and the Asia Pacific and India ( $R^2 = 0.30$ ). These results suggest that  $GC_{2 \times 2.5-200\%}$  is partially able to isolate NEE anomalies on continental scales that are separate from the large-scale ENSO-induced variability, and suggests that GOSAT flux inversions can be used to examine continental scale flux anomalies in the tropics. We note, however, that the the agreement in NEE IAV between  $GC_{2 \times 2.5-200\%}$  and FLUXCOM is not as strong in Africa and the Middle East.

### 3.2 Northern extratropics

10 In the northern extratropics, the observational coverage of GOSAT is highly seasonal and so we limit our analysis of anomalies in the northern extratropics to the summer (JJA), when observational coverage is the best (Liu et al., 2014; Byrne et al., 2017). Fig. 3 shows the anomalies for the proxies, FLUXCOM NEE, and  $GC_{2 \times 2.5-200\%}$  NEE across the northern hemisphere for JJA 2010–2013. The proxies and FLUXCOM generally show high coherence in anomalies. Events for which FLUXCOM NEE gives enhanced emission to the atmosphere also show reduced SIF, increased scPDSI, and increased  $T_{soil}$ . We have highlighted (with boxes) major climate anomalies over this time period: the 2010 Russian heat wave, the 2011 drought in Mexico and southern USA, the 2012 North American drought, and the 2013 California drought.  $GC_{2 \times 2.5-200\%}$  NEE indicates positive anomalies for all of these major events, suggesting that the inversion can recover sub-continental NEE IAV. However, there are also instances where the inversion and proxies tend to disagree. For example, in 2010,  $GC_{2 \times 2.5-200\%}$  indicates a positive anomaly in North America, whereas, the proxies indicate near neutral or negative anomalies.

20 To examine agreement with the proxies on regional-scale sub-continental regions, we have aggregated the inversions, the TBMs, proxies, and FLUXCOM into the five extratropical subcontinental regions shown in Fig. 1. The JJA anomalies in these regions over 2010–2013 provide 20 data points. We performed a linear regression of these anomalies against the proxies and FLUXCOM. Table 3 shows the slope and  $R^2$  values of the regressions. For the GOSAT inversions, the  $2^\circ \times 2.5^\circ$  and  $4^\circ \times 5^\circ$  with no prior NEE IAV show the closest agreement with FLUXCOM NEE and  $T_{soil}$  ( $0.49 \leq R^2 \leq 0.65$ ), while the inversions with prior NEE IAV show substantially poorer agreement ( $0.15 \leq R^2 \leq 0.36$ ). This is a larger difference between the inversions with and without prior NEE IAV than was found for the tropics (see Sec. 4.4.3). The inversions with NEE IAV also give a smaller slope indicating a smaller magnitude of NEE IAV, which is the opposite of what was found in the tropics. Comparing the inversions without prior NEE IAV, tight prior errors give  $0.57 \leq R^2 \leq 0.65$ , whereas loose prior constraints give  $0.49 \leq R^2 \leq 0.62$ . As with the tropics, the inversions with looser prior constraints give larger slopes, suggesting larger  
30 NEE IAV.

Comparing the other inversions, the GOSAT L4 product shows agreement with FLUXCOM NEE ( $R^2 = 0.33$ ) and  $T_{soil}$  ( $R^2 = 0.43$ ). CT2016 shows poor agreement with all proxies, indicating that this inversion is unable to isolate zonally asymmetric fluxes in the northern extratropics, which is surprising given the high sensitivity of the surface  $CO_2$  network to northern extratropical surface fluxes (Byrne et al., 2017). However, consistent with this result, Polavarapu et al. (2018) show that flux in-

versions assimilating measurements from the surface network are less able to recover zonally asymmetric flux signals than flux inversions assimilating GOSAT measurements. CT2016 also includes prior NEE IAV in the inversion, which may negatively impact the posterior NEE IAV, based on the GEOS-Chem inversion results.

For the TBMs, VISIT shows close agreement with FLUXCOM NEE, scPDSI, and  $T_{soil}$  anomalies and to a lesser extent SIF anomalies. This is notable as VISIT generally showed poor agreement with the proxies in the tropics. JULES shows close agreement with  $T_{soil}$  anomalies and some agreement with the other proxies. CASA GFED 4.1 shows good agreement with SIF anomalies, but comparatively poorer agreement with the other proxies. CASA GFED CMS shows some agreement with SIF anomalies, but little agreement with the other proxies.

### 3.3 Observing system simulation experiments

~~We performed a series of OSSE experiments. Strong correlations between the GOSAT flux inversions and proxies/FLUXCOM provide evidence that the GOSAT flux inversions give realistic constraints on NEE. However, the absence of strong correlations does not imply that the GOSAT flux inversions are not constraining IAV as there could be other causes (such as lagged effects within ecosystems) that can explain the absence of correlations. Therefore,~~ to investigate the minimum spatial scales that can be constrained by GOSAT observations, we performed a series of OSSE experiments. In these experiments pseudo-observations were assimilated from a GEOS-Chem forward model run which had JULES NEE fluxes prescribed. See Sect. 2.3.4 for additional details of the OSSE setup.

#### 3.3.1 Tropics

Strong correlations are obtained between the posterior and true anomalies for all OSSEs on sub-continental regional and larger scales, suggesting that sub-continental regions are the minimum scales that can be constrained by GOSAT measurements. Figure 4 shows the mean correlation coefficient between the posterior and true NEE anomalies in the tropics over a range of scales. ~~The results are highly reminiscent of the results between the GOSAT inversion and the proxies. The mean correlation between the posterior and true NEE anomalies increases with aggregation for all OSSEs. Strong correlations are obtained for all OSSEs on regional and larger scales.~~ The inversion setup ~~also~~ has an impact on the correlations between the posterior and true NEE IAV. The largest differences between OSSEs are obtained on sub-continental regional and continental scales. On these scales, OSSE<sub>JULES-100%</sub> has the largest correlation. ~~This suggests that having a climatological seasonal cycle close to the true NEE IAV is important for recovering NEE IAV in the tropics.~~ The inclusion of prior NEE IAV (OSSE<sub>CT2016-100%-IAV</sub>) does not appear to ~~significantly~~ substantially degrade the correlation relative to a prior NEE without IAV (OSSE<sub>CT2016-100%</sub>). ~~In fact, inclusion of prior NEE IAV actually improves the correlations (except for continental scale),~~ in contrast to what was found with the real data GOSAT inversions. The prior error constraints generally have a large influence on the correlation with the true NEE IAV. Loose prior constraints give better agreement for all OSSEs, consistent with the GOSAT inversions.

On the scale of the entire tropics, we performed linear regressions between the posterior and true anomalies, which are shown in Table 4. For all regressions, the magnitude of IAV in the posterior fluxes is less than the true NEE IAV (slope of 0.42–0.75). This suggests that the inversions do not recover the full magnitude of NEE IAV. In addition to comparing

posterior and true anomalies, we examine the similarities in posterior anomalies between OSSEs. The right column of Table 4 shows the results of linear regressions between posterior and OSSE<sub>JULES-100%</sub> NEE anomalies. The OSSEs without prior NEE IAV show better agreement with OSSE<sub>JULES-100%</sub> posterior anomalies than the true anomalies. This suggests that the assimilation of pseudo-data is introducing NEE anomalies in a similar way for all OSSEs and recovering the true NEE IAV is primarily limited by the observational coverage rather than the inversion setup. ~~However, differences between the OSSEs and true NEE IAV may also be due to systematic biases introduced due to factors such as uneven observational coverage (Liu et al., 2014; Byrne et al., 2017).~~ Overall, these OSSEs suggest that the real data flux inversions underestimate IAV in NEE by 42–75%, and suggest that sub-continental regions are the minimum scale for which IAV can be recovered from GOSAT measurements.

~~We examine the continental-scale anomalies in detail for OSSE<sub>JULES-100%</sub>, OSSE<sub>CT2016-100%</sub>, and OSSE<sub>CT2016-100%-IAV</sub> in Figure ??, which shows the timeseries of continental-scale flux anomalies in the tropics for the OSSEs. The correlation between the OSSEs and true anomalies improves after performing a three-month running mean, consistent with the GOSAT inversion results. Strong correlations between the OSSEs and true NEE IAV are obtained after removing the mean tropical signal (using equation ??). These results provide further evidence that GOSAT inversions can largely recover continental-scale flux anomalies in the tropics.~~

### 3.3.2 Northern extratropics

The posterior NEE anomalies are generally correlated with the truth ( $0.15 < R^2 < 0.48$ ) but give slopes less than one, indicating that the OSSEs are not recovering the full magnitude of NEE IAV. Table 4 shows the slope and  $R^2$  for linear regressions of ~~flux~~ JJA NEE anomalies from the OSSEs against the true NEE IAV on sub-continental regions in the northern extratropics during JJA. ~~In all cases the slope is less than one, indicating that the OSSEs are not recovering the full magnitude of NEE IAV. The  $R^2$  values are less than between the GOSAT inversions and proxies. This may be due to the fact that temporal anomalies in JULES NEE are highly variable month-to-month and may have a shorter temporal correlation length scales than the true anomalies.~~ Comparing the different OSSE setups, ~~the~~ OSSE<sub>CT2016-100%-IAV</sub> performs substantially worse than the OSSEs with no prior NEE IAV ( $R^2 = 0.15$  versus  $R^2 = 0.30-0.48$ ). This is consistent with comparisons between GOSAT inversions and proxies, and suggests that employing prior NEE IAV in the northern extratropics degrades posterior NEE IAV on sub-continental scales during JJA. OSSEs with tighter prior constraints give larger  $R^2$ , consistent with the GOSAT inversions. OSSEs with JULES mean seasonal cycle also agree better with the true NEE IAV than those which employ the mean seasonal cycle from CT2016.

## 4 Discussion

### 4.1 Scales constrained by GOSAT $X_{CO_2}$

We investigated the agreement between monthly posterior NEE anomalies and proxy/FLUXCOM anomalies over a range of spatial scales. The results showed that the agreement between the inversions and the proxies/FLUXCOM were scale-dependent.



Here we synthesize these results and discuss the ability of GOSAT flux inversions to recover IAV in NEE over the range of scales examined in this study.

In the tropics, we find strong evidence that the GOSAT inversions are recovering the timing of NEE anomalies on the scale of the entire tropics, however, there is less agreement on the magnitude of monthly NEE anomalies over the tropics. The OSSEs suggest that GOSAT flux inversions underestimate the magnitude of NEE anomalies by about 50%, depending on the inversion setup. On continental scales in the tropics, strong correlations with FLUXCOM NEE and  $T_{soil}$  anomalies suggest that GOSAT flux inversion can constrain monthly NEE anomalies. On sub-continental regional scales in the tropics, the results were more ambiguous. The GOSAT inversions generally showed good agreement with FLUXCOM NEE IAV on sub-continental regional scales, but only marginal agreement with  $T_{soil}$ . The OSSEs also indicate marginal ability to recover sub-continental regional scale fluxes. From these results, we caution against making conclusions about NEE IAV on sub-continental regional scales in the tropics using GOSAT flux inversions without corroborating evidence. On smaller scales, correlations do not meet the threshold of  $P < 0.05$ .

In the northern extratropics during JJA, the results of this study suggest that sub-continental regional-scale constraints are possible. We found that large NEE anomalies due to major climate events are recovered in the inversion for  $GC_{2 \times 2.5-200\%}$  (Fig 3), while linear regressions showed close agreement for the GOSAT flux inversions with FLUXCOM NEE and  $T_{soil}$ . However, we also found evidence that the posterior NEE IAV was sensitive to the inversion setup. The inversion analyses with prior NEE IAV ( $GC_{4 \times 5-44\%-IAV}$ ,  $GC_{4 \times 5-100\%-IAV}$ , and GOSAT L4) showed weaker agreement with the proxies relative to the inversions without prior NEE IAV. Similarly, the OSSEs showed prior NEE IAV reduced agreement between the posterior and the “true” NEE IAV in the northern extratropics during JJA.

## 4.2 Influence of ENSO

ENSO is the primary driver of large scale NEE IAV in the tropics, therefore, it is useful to consider correlations between the inversions/TBMs and proxies/FLUXCOM within the context of ENSO variability. Here, we contrast correlations between the inversions and proxies to correlations between the NINO 3.4 index and proxies. Figure 2 shows correlations between the NINO 3.4 index and proxies over the range of spatial scales. Correlations between the proxies and the NINO 3.4 index increase with aggregation, as expected since the NINO 3.4 index reflects the large scale ENSO-driven variability in the tropics. The GEOS-Chem GOSAT flux inversions show stronger correlations with the proxies than the NINO 3.4 index on sub-continental regions and at larger scales for correlations with FLUXCOM NEE and at continental and larger scales for  $T_{soil}$ . For the GOSAT L4 inversion, the correlation only reaches this threshold for  $T_{soil}$  at the largest aggregation scale. These results suggest that GOSAT observations provide more detailed flux information on continental and larger scales than the ENSO index. Thus, these data could provide insights into differences in the carbon dynamics between ENSO events, and supports the ability of space-based  $X_{CO_2}$  to inform the continental-scale responses to ENSO variability as presented by Liu et al. (2017). Additional comparisons between posterior NEE fluxes and the NINO 3.4 index are provided in Sec. S1 of the supplementary materials.

### 4.3 Implications of correlations between flux inversions and proxies

The results of this study show varying degrees of agreement between anomalies in GOSAT flux inversions and anomalies in proxies and FLUXCOM. We consistently find that  $T_{soil}$  and FLUXCOM NEE show the strongest agreement with the flux inversions, whereas scPDSI and SIF show weaker agreement. In this section, we discuss agreement between the proxies and  
5 flux inversions in detail, and their implications.

#### 4.3.1 Agreement with $T_{soil}$ and scPDSI

The results show high consistency in the timing of anomalies between  $T_{soil}$  and GOSAT flux inversions on continental and larger scales in the tropics, and on sub-continental scales in the northern extratropics during JJA. These results indicate that  $T_{soil}$  is a useful proxy for corroborating NEE IAV in flux inversions in both the tropics and northern extratropics. ~~Linear regressions between GOSAT flux inversion and scPDSI IAV indicate moderate agreement on the scale of the entire tropics ( $R^2 \leq 0.27$ ) and on sub-continental scales in the northern extratropics ( $R^2 \leq 0.29$ ).~~ The GOSAT flux inversion NEE IAV consistently shows closer agreement with  $T_{soil}$  anomalies than with scPDSI in both the tropics and northern extratropics. This is consistent with previous research that has ~~mostly~~ shown that NEE IAV is most closely related to temperature anomalies on large scales (Wang et al., 2013; Jung et al., 2017).

15 Although the results of this study indicate that  $T_{soil}$  is a useful metric for corroborating NEE IAV in flux inversions, inferring the sensitivity of NEE anomalies to temperature anomalies directly is not advised for the fits given in Tables 2 and 3. This is because a number of factors have not been considered in this analysis. One factor is that temperature anomalies are also correlated with moisture and biomass burning anomalies. Keppel-Aleks et al. (2014) show that accounting for these covariances results in reduced sensitivity of NEE anomalies to temperature anomalies. A second factor is that the relationship between NEE  
20 anomalies and temperature and moisture anomalies is variable, depending on large scale climate modes. For example, Fang et al. (2017) show that either temperature or precipitation anomalies can be the primary driver NEE anomalies based on ENSO phase. A third factor is that the impact of temperature and moisture on NEE anomalies may be lagged (Braswell et al., 1997). Ecosystems can take a months to years to recover from droughts (Frank et al., 2015; Schwalm et al., 2017; Sippel et al., 2018). Baldocchi et al. (2018) found that ~~flux~~ NEE anomalies at number of FLUXNET sites are negatively correlated with themselves  
25 after a one-year lag, implying a highly oscillatory behavior in the net carbon fluxes from year to year.

This leaves many opportunities for future work to further investigate the relationship between NEE anomalies and climate variability in more detail. A further limit to the comparisons of flux inversions with  $T_{soil}$  and scPDSI anomalies in the tropics is that we do not distinguish between seasons. The relationship between NEE,  $T_{soil}$  and scPDSI anomalies likely have substantial seasonal differences (Rödenbeck et al., 2018). We encourage future studies to examine the seasonally-dependent relationships  
30 using longer flux inversions, as well as studies which investigate lagged correlations and climate mode relationships between inversion NEE anomalies and temperature and water availability anomalies.

### 4.3.2 Agreement with SIF

It is notable that correlations with SIF are weaker than those with the other proxies. Linear regressions indicate that SIF anomalies show some correspondence to GOSAT flux inversion anomalies on sub-continental scales in the northern extratropics during JJA ( $0.14 \leq R^2 \leq 0.27$ ), but little agreement is found in the tropics ( $R^2 \leq 0.05$ ). These results are not all that surprising, as it is not clear that one should expect SIF and NEE to be highly correlated, since SIF is a proxy of GPP rather than NEE. It has previously been shown that NEE and GPP anomalies are only moderately correlated (Baldocchi et al., 2018). However, we also note that spurious trends have been found in the [GOME-2 SIF](#) observations (Zhang et al., 2018), which could impact the calculated anomalies. Furthermore, due to GOME-2's large field of view, clouds are almost always present for measurements in the tropics. ~~To test if the~~ [We compared anomalies in GOME-2 SIF anomalies used here, we examined the correlation between with FLUXCOM MARS GPP and SIF anomalies \(Fig. ??\). Spatially heterogeneous agreement is found between the two datasets, with the closest agreement occurring over semi-arid regions. However, anomalies and found that](#) correlations are generally positive over the majority of the globe, suggesting that IAV from GOME-2 SIF is reliable ~~-(Fig. S4). However, the correlation are also spatially heterogeneous, with the closest agreement occurring over semi-arid regions.~~

### 4.3.3 Agreement with FLUXCOM NEE

The GEOS-Chem GOSAT flux inversions with no prior NEE IAV showed close agreement with FLUXCOM NEE anomalies in the tropics on [sub-continental](#) regional and larger scales, and in the northern extratropics on [sub-continental](#) regional scales during JJA. This is a remarkable finding as these data-driven estimates of NEE IAV are independent, and agreement between the two estimates provides a strong indication that the results are robust. Therefore, comparisons with FLUXCOM NEE may provide a method for corroborating results from flux inversion studies. However, it should be noted that the net annual NEE fluxes produced by FLUXCOM are quite unrealistic (Tramontana et al., 2016; Jung et al., 2017), with annual net draw-down by the biosphere of 18–28 PgC yr<sup>-1</sup>.

It may also be possible to evaluate the magnitude of NEE IAV in FLUXCOM NEE through comparisons with flux inversions. Here we compare the magnitude of NEE IAV between the GOSAT flux inversions and FLUXCOM. The slope of the linear regression between the inversions indicates the relative magnitude of the inversion and FLUXCOM NEE anomalies. Over the entire tropics, the GOSAT inversions give slopes of 1.03–2.10 (mean of 1.56), suggesting that the magnitude of NEE anomalies are underestimated by FLUXCOM NEE. For JJA in the northern extratropics, the GOSAT inversions give slopes of 0.79–1.59 (mean of 1.31), again suggesting that the magnitude of NEE anomalies are underestimated by FLUXCOM. Furthermore, the OSSEs suggested that the inversions do not recover the full magnitude of NEE IAV, providing further evidence that FLUXCOM underestimates the magnitude of NEE IAV. This result is consistent with previous studies which indicate that FLUXCOM underestimate the magnitude of NEE IAV (Jung et al., 2011, 2017).

## 4.4 Scales constrained

We investigated the agreement between monthly anomalies in flux inversions and proxies/FLUXCOM over a range of spatial scales in the tropics. The results showed that the agreement between the inversions and the proxies/FLUXCOM were scale-dependent, which was corroborated by OSSEs. Here we synthesize these results and discuss the ability of GOSAT flux inversions to recover IAV in NEE over the range of scales examined in this study.

5 The results provide strong evidence that GOSAT flux inversion can constrain monthly flux anomalies on the scale of the entire tropics. All of the GEOS-Chem GOSAT flux inversions obtained  $R^2 \geq 0.55$  for linear regressions with  $T_{soil}$ , and  $R^2 \geq 0.51$  with FLUXCOM NEE. The OSSEs provide further evidence that the true NEE anomalies could be recovered, as linear regressions between the posterior and true anomalies give  $R^2 \geq 0.53$ . These results provide strong evidence that the GOSAT inversions are recovering the timing of tropical NEE anomalies, however, there is less agreement on the magnitude  
10 of flux anomalies over the tropics. The OSSEs indicate that GOSAT flux inversions can recover 42–68% of the magnitude of NEE anomalies, depending on the inversion setup.

On continental scales in the tropics, the results suggest that GOSAT flux inversion can constrain monthly flux anomalies. The GEOS-Chem inversions show good agreement with FLUXCOM NEE and  $T_{soil}$  anomalies. However, the agreement between the inversions and proxies/FLUXCOM on this scale is strongly influenced by the large-scale ENSO anomalies. We isolated  
15 the continental-scale anomalies by subtracting a mean tropical anomaly for GC<sub>2x2.5-200%</sub> and FLUXCOM (Fig. ??), and for the OSSEs (Fig. ??). We found that the anomalies were still correlated after removing the mean tropical signal, suggesting that the continental-scale anomalies are largely recovered in the inversions. However, we also found that the inversion setup can have a significant influence on posterior anomalies on continental scales. The strongest correlations between the proxies and inversions were obtained with higher resolution, looser prior constraints, and no prior NEE IAV. Similarly, OSSEs showed  
20 the best agreement with the true NEE IAV when looser prior constraints were employed, but suggested that the presence of prior NEE IAV generally improved agreement with the true NEE IAV. The OSSEs also showed that correlations with the true NEE IAV were improved on continental scales when the prior mean seasonal cycle was closer to the true NEE IAV. Overall, these results suggest that GOSAT observations contain information on continental-scale NEE anomalies in the tropics; however, recovering the correct NEE IAV from these observations may be sensitive to the flux inversion setup. Furthermore,  
25 the magnitude of NEE anomalies are likely underestimated.

On regional scales in the tropics, the results were more ambiguous. The GOSAT inversions generally showed good agreement with FLUXCOM NEE IAV on regional scales, but only marginal agreement with  $T_{soil}$ . The OSSEs also indicate marginal ability to recover regional scale fluxes. From these results, we caution against making conclusions about NEE IAV on regional  
30 scales in the tropics using GOSAT flux inversions without corroborating evidence. On smaller scales, correlations do not meet the threshold of  $P < 0.05$ .

In the northern extratropics during JJA, the results of this study suggest that regional-scale constraints are possible. We found that large flux anomalies due to major climate events are recovered in the inversion for GC<sub>2x2.5-200%</sub> (Fig 3), while linear regressions showed close agreement for the GOSAT flux inversions with FLUXCOM NEE and  $T_{soil}$ . However, we also found evidence that the posterior NEE IAV was sensitive to the inversion setup. The inversion analyses with prior NEE IAV  
35 (GC<sub>4x5-44%-IAV</sub>, GC<sub>4x5-100%-IAV</sub>, and GOSAT L4) showed weaker agreement with the proxies relative to the inversions

~~without prior NEE IAV. Similarly, the OSSEs showed prior NEE IAV reduced agreement between the posterior and the “true” NEE IAV in the northern extratropics during JJA.~~

#### 4.4 Influence of the inversion configuration

##### 4.4.1 Model horizontal resolution

5 The results of this study indicate that the spatial resolution of the model used in the inversion analysis ( $2^\circ \times 2.5^\circ$  or  $4^\circ \times 5^\circ$ ) has a relatively minor impact on posterior NEE anomalies. This somewhat surprising since recent studies (Yu et al., 2018; Stanevich, 2018) have shown significant transport differences for different resolution versions of GEOS-Chem. Also, Deng et al. (2015) showed that there are large biases in  $\text{CO}_2$  in the upper troposphere and lower stratosphere in GEOS-Chem that impact inferred flux estimates. It is possible that, although the model transport errors influence the flux estimates, the resolution-  
10 dependent transport processes are not sensitive to NEE IAV for the time period considered here. ~~It could also be related to the information content of GOSAT observations. As we have shown~~ However, both model resolutions examined in this study ; ~~GOSAT observations only constrain NEE IAV on regional and larger scales. If transport errors have the largest impact on smaller scales, it may explain why model resolution did not have a major impact on our results.~~ are quite coarse, and horizontal resolution becomes increasingly important at smaller spatial scales (Agustí-Panareda et al., 2019).

##### 15 4.4.2 Prior error covariances

~~All of the GEOS-Chem inversions were performed with tight (44% for  $4^\circ \times 5^\circ$  and 66% for  $2^\circ \times 2.5^\circ$ ) and loose (100% for  $4^\circ \times 5^\circ$  and 200% for  $2^\circ \times 2.5^\circ$ ) prior error covariances. The~~ The magnitude of the prior error covariances generally had a significant impact on the posterior NEE IAV. In the tropics, inversions with loose prior constraints gave larger correlations with  $T_{soil}$  and FLUXCOM NEE on ~~regional and continental scales. Similarly, for the OSSEs, looser prior constraints gave larger~~  
20 ~~correlations with the true NEE IAV on~~ sub-continental regional and continental scales. This suggests that the information content of the GOSAT observations is sufficiently large in the tropics that prior error covariances of 100% for  $4^\circ \times 5^\circ$  or 200% for  $2^\circ \times 2.5^\circ$  can be applied without degrading the posterior results by over-fitting the observations.

In the northern extratropics, the inversions with tighter prior constraints gave larger correlations with  $T_{soil}$  and FLUXCOM NEE on sub-continental regional and continental scales. ~~Similarly, tight prior constraints gave larger correlations with the true~~  
25 ~~NEE IAV for the OSSEs, suggesting that loose prior constraints result in overfitting of the GOSAT observations, which degrades the agreement with proxies.~~ These results are the opposite of what was found for the tropics, and suggests that tighter error constraints (as a percentage of NEE) should be applied in the northern extratropics than in the tropics. ~~These results suggest that, when the prior error covariances are loose in the northern extratropics, the inversion over-fits the GOSAT observations which degrades the agreement with proxies (or the true NEE IAV for OSSEs).~~

30 The largest impact of varying the prior error covariances is in the magnitude of posterior NEE IAV. When loose prior constraints are applied the magnitude of NEE anomalies increases by 30–80% (15–30% for OSSEs) in the tropics and 5–60% (0–30% for OSSEs) in the northern extratropics. These results imply that care should taken when making conclusions about

the magnitude of NEE anomalies from this analysis. Based on the OSSEs, it seems likely that the inversions underestimate the magnitude of NEE IAV on all scales, but the inversions with looser prior constraints result in a more realistic magnitude of NEE IAV. This suggests that there is a trade off between obtaining a more realistic magnitude of IAV using looser constraints and obtaining more realistic timing of anomalies with tighter prior constraints.

#### 5 4.4.3 Prior fluxes

~~We investigated the influence of prior NEE IAV on posterior NEE anomalies in flux inversions by performing inversions with prior NEE IAV ( $GC_{4 \times 5} = 100\% - IAV$  and  $GC_{4 \times 5} = 44\% - IAV$ ) and without prior NEE IAV ( $GC_{4 \times 5} = 100\%$  and  $GC_{4 \times 5} = 44\%$ ), as well as OSSEs with and without prior NEE IAV. In the tropics, the impact of prior NEE IAV is generally small. For the GOSAT inversions, the presence of prior NEE IAV degrades agreement with the proxies on all scales. In the OSSEs, the presence of prior NEE IAV degrades the agreement with the true NEE IAV on continental scales, but improves agreement on regional scales and over the entire tropics. In~~

~~The presence of IAV in prior NEE resulted in reduced correlations between NEE and the proxies in the northern extratropics, the presence while the impact of prior NEE IAV has a large negative impact on agreement with proxies for GOSAT inversions and on agreement with the true NEE IAV in the OSSEs.~~

15 ~~Why does the was generally small in the tropics. The presence of prior NEE IAV degrade likely degrades the posterior NEE IAV for many of these inversions? Presumably, the reason is related due to the fact that the observations under-constrain NEE IAV, such that the prior NEE IAV strongly influences the spatiotemporal distribution of IAV in the posterior NEE. ~~To investigate this, we examined how closely~~ In fact, the posterior NEE IAV resembles the prior NEE IAV. Figure ?? shows the agreement between the posterior and prior NEE IAV for  $GC_{4 \times 5} = 100\% - IAV$  in the tropics and northern extratropics. Posterior~~

20 ~~NEE IAV is~~ is generally more strongly correlated with IAV in the prior NEE, particularly on smaller scales. The fact that correlations between the prior and posterior NEE IAV are strong at  $4^\circ \times 5^\circ$  and  $8^\circ \times 10^\circ$  is not surprising, as the NEE fluxes are strongly under-constrained at these spatial scales. However, the correlation with the prior NEE IAV is substantially larger than with FLUXCOM on regional ( $R^2 = 0.55$  versus  $R^2 = 0.15$ ) and continental ( $R^2 = 0.46$  versus  $R^2 = 0.26$ ) scales as well. This suggests that NEE IAV is still under-constrained even on continental scales the prior IAV than the proxies/FLUXCOM.

25 Only on the scale of the entire tropics is the correlation with the prior NEE ( $R^2 = 0.42$ ) less than with the proxies ( $R^2 = 0.61$  for FLUXCOM NEE and  $R^2 = 0.56$  for  $T_{soil}$ ), indicating that the observations are influencing the posterior NEE IAV more than the prior NEE IAV. ~~These results~~ (Sec. S2). This suggests that the impact of prior NEE IAV on the inversion is likely strongly dependent on how well the prior NEE IAV reflects the true NEE IAV. Realistic prior NEE IAV would likely improve the posterior NEE IAV, conversely, unrealistic prior NEE IAV will degrade the posterior NEE IAV. This implies that the realism

30 of the prior NEE IAV should be investigated before including it in an inversion analysis. If the objective of the experiment is to examine the timing of posterior NEE IAV introduced through the assimilation of observations, then we recommend that annually-repeating prior fluxes be used to investigate NEE IAV. ~~However, a trade-off in using annually-repeating prior fluxes is that the magnitude of NEE IAV will likely be significantly underestimated.~~

~~We also investigated the The impact of the prior mean seasonal cycle on posterior NEE IAV. ~~We performed~~ was investigated in a series of OSSEs ~~to examine the impact of the mean seasonal cycle of the prior fluxes on the inversion and~~. We found that correlations with the true NEE IAV were ~~significantly~~ improved on continental scales when the mean seasonal cycle was closer to the true NEE IAV. ~~In particular, OSSE<sub>CT2016=100%</sub> gives much weaker correlations with the true NEE IAV than~~ OSSE<sub>JULES=100%</sub> after removing the mean tropical signal (Fig. ??). These results suggest that it is important to use prior fluxes with a realistic seasonal cycle to recover IAV in NEE from GOSAT observations.~~

## 5 Conclusions

In this study, we examined the constraints on ~~interannual anomalies in NEE~~ NEE IAV provided by GOSAT observations by performing a series of flux inversions. We addressed ~~three~~ two main objectives in this analysis. The first objective was to quantify the agreement between GOSAT flux inversions and flux proxies, which are associated with IAV in the terrestrial carbon cycle, and FLUXCOM NEE. We found ~~strong correlations~~ ( $P < 0.05$ ) that posterior NEE anomalies were strongly correlated with FLUXCOM NEE and  $T_{soil}$  anomalies,  $R > R_{NINO3.4}$  with FLUXCOM NEE and  $T_{soil}$  in the tropics on continental and larger scales, and in the northern extratropics on sub-continental scales during the summer ( $R^2 > 0.49$ ), when there is no prior NEE IAV. These results demonstrate that both FLUXCOM NEE and  $T_{soil}$  can be useful tools for corroborating flux inversion results. ~~We found flux anomalies from GOSAT inversions were less correlated with scPDSI and SIF. For scPDSI we found some agreement on the scale of the entire tropics ( $R^2 \leq 0.27$ ) and on sub-continental scales in the northern extratropics ( $R^2 \leq 0.29$ ). For SIF, there was some agreement on sub-continental scales in the northern extratropics during JJA ( $0.14 \leq R^2 \leq 0.27$ ), however, little agreement was found in the tropics ( $R^2 \leq 0.05$ ).~~

The second objective was to determine the spatial scales over which the GOSAT inversion constrain flux anomalies. ~~In the tropics, we found that continental and larger scale flux anomalies can be captured in GOSAT inversions. This conclusion is supported by strong agreement but gave weaker correlations with scPDSI and SIF. Strong correlations ( $P < 0.05$ ,  $R > R_{NINO3.4}$ ) with  $T_{soil}$  and FLUXCOM NEE, and a series of OSSEs which showed that the true NEE IAV can be largely recovered on these scales. On regional~~ we were found with FLUXCOM NEE and  $T_{soil}$  on continental and larger scales in the tropics, the suggesting that GOSAT flux inversions showed some agreement with the proxies and FLUXCOM, but the OSSEs indicated that the GOSAT observations likely underestimate can capture NEE IAV on these and smaller scales. spatial scales. In the northern extratropics, ~~we found that flux anomalies are recovered by GOSAT flux inversions on sub-continental regions during JJA. Strong agreement was found~~ JJA posterior NEE anomalies were found to show strong agreement with anomalies in  $T_{soil}$  ( $0.57 \leq R^2 \leq 0.65$ ) and FLUXCOM NEE ( $0.49 \leq R^2 \leq 0.65$ ) on sub-continental regions, when no prior NEE IAV is used. ~~OSSEs supported these findings, indicating that GOSAT observations can recover regional scale flux anomalies in the northern extratropics during JJA.~~

The ~~third~~ second objective was to quantify the sensitivity of ~~the results from the first two objectives~~ correlations between posterior NEE anomalies and proxy/FLUXCOM anomalies to the inversion setup. We found that the agreement between the flux inversions and proxies can be sensitive to the inversion setup. Posterior ~~flux~~ NEE anomalies were most sensitive to the

prior fluxes and error covariances. In general, the The inclusion of prior NEE IAV from CT2016 in the inversion degraded the agreement with FLUXCOM NEE and  $T_{soil}$ , particularly in the extratropics. This result was supported by the OSSEs in the northern extratropics but not in the tropics. We compared the impact of the mean seasonal cycle on the posterior NEE IAV by performing OSSEs and found that having a prior climatological seasonal cycle that was close to the true NEE IAV improved posterior NEE anomalies on continental scales in the tropics. The prior error constraints also had a significant impact on the results. We found that looser constraints in the tropics gave better agreement with the proxies, while tighter constraints in the northern extratropics gave better agreement with the proxies. The magnitude of the prior constraints had a large had a substantial impact on the magnitude of NEE anomalies. Also, the posterior NEE anomalies, and OSSEs showed that the magnitude of NEE anomalies are underestimated even with loose prior constraints. These results indicate that the prior fluxes and error covariances need to be carefully considered. The inclusion of prior NEE IAV is an important factor to consider. Including prior NEE IAV may produce a more realistic magnitude of NEE IAV in posterior fluxes but could also prior NEE IAV can degrade the correlation between the posterior and true NEE IAV. If Therefore, if prior NEE IAV are included in future inversions, attempts should be made to test the realism of the prior NEE IAV. If the objective of the experiment is to examine the timing of posterior NEE IAV introduced through the assimilation of observations, then we recommend that annually-repeating prior fluxes be used to investigate NEE IAV. The mean seasonal cycle of the prior NEE fluxes is also an important factor in the inversion, and the realism the seasonal cycle should be evaluated before performing the inversion. The results also indicate that defining the prior error covariance to be a fraction of the prior flux may produce either overfitting of GOSAT data in the northern extratropics or underfitting of the data in the tropics.

The magnitude of prior error covariances had a large impact on the magnitude of the posterior IAV, but the amplitude of NEE IAV is likely underestimated even if loose prior constraints are applied. Although not addressed in this study, correlated errors between GOSAT observations may introduce structures in the posterior NEE estimates, thus we recommend future work address the possibility of prescribing non-diagonal terms in the observational error covariance matrix.

Overall, our results show that  $T_{soil}$  and FLUXCOM NEE are useful for evaluating NEE IAV in flux inversions. Furthermore, comparisons with the anomalies in  $T_{soil}$  and FLUXCOM NEE suggest that GOSAT flux inversions are useful tools for constraining IAV in NEE on continental and larger scales in the tropics, and on regional scales in the northern extratropics during JJA. The results of this study suggest that GOSAT measurements provide a useful constraint on IAV in the carbon cycle. Further study of the relationship between GOSAT-constrained NEE and environmental variables is merited given results discussed here. In particular, the mechanisms driving these co-variations should be further investigated. Lagged relationships between GOSAT-constrained NEE and environmental variables should also be investigated. Future research could also investigate differences in IAV between GOSAT-constrained NEE and that produced by TBMs. Given the better agreement with the proxies, GOSAT-constrained NEE IAV may provide a tool for evaluating the TBM-simulated NEE IAV in the future.

*Competing interests.* No competing interests are present



*Acknowledgements.* Funding for this work was provided by the Canadian Space Agency, NSERC, and Environment and Climate Change Canada. CarbonTracker CT2016 results were provided by NOAA ESRL, Boulder, Colorado, USA from the website at <http://carbontracker.noaa.gov>. CASA GFED 4.1 and CASA CMS NEE fluxes were also downloaded from the CT2016 website. The GOSAT L4 product and VISIT NEE were downloaded from the GOSAT Data Archive Service (<https://data2.gosat.nies.go.jp>). Dai Global Palmer Drought Severity Index was downloaded from the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (<https://doi.org/10.5065/D6QF8R93>), accessed 15 Aug 2017. NASA GOME-2 SIF products were obtained from Aura Validation Data Center [<http://avdc.gsfc.nasa.gov>]. FLUXCOM products were obtained from the Data Portal of the Max Planck Institute for Biochemistry [<https://www.bgc-jena.mpg.de>]. MERRA-2 products were downloaded from MDISC [<https://disc.sci.gsfc.nasa.gov>], managed by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). The GEOS-Chem forward and adjoint models are freely available to the public. Instructions for downloading and running the models can be found at <http://wiki.geos-chem.org/>. ACOS GOSAT lite files were obtained from the CO<sub>2</sub> Virtual Science Data Environment [<https://co2.jpl.nasa.gov/#mission=ACOS>]. The SST anomalies were downloaded from the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) website (<https://www.esrl.noaa.gov>). We thank Martin Jung, Andy Jacobson, Joanna Joiner, Randy Kawa, and Jim Collatz for helpful comments on this manuscript.

## References

- Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz Sabater, J., Barré, J., Curcoll, R., Engelen, R., Langerock, B., Law, R. M., Loh, Z., Morguí, J. A., Parrington, M., Peuch, V.-H., Ramonet, M., Roehl, C., Vermeulen, A. T., Warneke, T., and Wunch, D.: Modelling CO<sub>2</sub> weather – why horizontal resolution matters, *Atmos. Chem. Phys.*, 19, 7347–7376, <https://doi.org/10.5194/acp-19-7347-2019>, <https://www.atmos-chem-phys.net/19/7347/2019/>, 2019.
- Andres, R., Boden, T., and Marland, G.: Monthly Fossil-Fuel CO<sub>2</sub> Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A, <https://doi.org/10.3334/CDIAC/ffe.MonthlyMass.2016>, 2016.
- Bacastow, R.: Modulation of atmospheric carbon dioxide by the Southern Oscillation, *Nature*, 261, 116–118, 1976.
- 10 Bacastow, R., Adams, J., Keeling, C., Moss, D., Whorf, T., and Wong, C.: Atmospheric carbon dioxide, the Southern Oscillation, and the weak 1975 El Niño, *Science*, 210, 66–68, 1980.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., et al.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem–scale carbon dioxide, water vapor, and energy flux densities, *B. Am. Meteorol. Soc.*, 82, 2415–2434, 2001.
- 15 Baldocchi, D., Chu, H., and Reichstein, M.: Inter-annual variability of net and gross ecosystem carbon fluxes: A review, *Agricultural and Forest Meteorology*, 249, 520–533, <https://doi.org/https://doi.org/10.1016/j.agrformet.2017.05.015>, <http://www.sciencedirect.com/science/article/pii/S0168192317301806>, 2018.
- Battle, M., Bender, M., Tans, P. P., White, J., Ellis, J., Conway, T., and Francey, R.: Global carbon sinks and their variability inferred from atmospheric O<sub>2</sub> and  $\delta^{13}\text{C}$ , *Science*, 287, 2467–2470, 2000.
- 20 Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al.: Redefine statistical significance, *Nat. Hum. Behav.*, 2, 6, 2018.
- Berry, J. and Bjorkman, O.: Photosynthetic response and adaptation to temperature in higher plants, *Annual Review of plant physiology*, 31, 491–543, 1980.
- Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P., and Tans, P. P.: Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, 290, 1342–1346, 2000.
- 25 Braswell, B., Schimel, D. S., Linder, E., and Moore, B.: The response of global terrestrial ecosystems to interannual temperature variability, *Science*, 278, 870–873, <https://doi.org/10.1126/science.278.5339.870>, 1997.
- Bruhwyler, L., Michalak, A., and Tans, P.: Spatial and temporal resolution of carbon flux estimates for 1983–2002, *Biogeosciences*, 8, 1309–1331, 2011.
- 30 Byrne, B., Jones, D. B. A., Strong, K., Zeng, Z.-C., Deng, F., and Liu, J.: Sensitivity of CO<sub>2</sub> Surface Flux Constraints to Observational Coverage, *J. Geophys. Res.-Atmos*, 112, 6672–6694, <https://doi.org/10.1002/2016JD026164>, 2017.
- Byrne, B., Wunch, D., Jones, D. B. A., Strong, K., Deng, F., Baker, I., Koehler, P., Frankenberg, C., Joiner, J., Arora, V. K., et al.: Evaluating GPP and respiration estimates over northern midlatitude ecosystems using solar induced fluorescence and atmospheric CO<sub>2</sub> measurements, *J. Geophys. Res.-Biogeo.*, 123, 1–22, <https://doi.org/10.1029/2018JG004472>, 2018.
- 35 Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., et al.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, 437, 529–533, 2005.

- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, pp. 465–570, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.015>, 2013.
- Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al.: The Joint UK Land Environment Simulator (JULES), model description-Part 2: carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, 4, 701–722, <https://doi.org/10.1029/93GB02725>, 2011.
- Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A., and Zhang, N.: Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory global air sampling network, *J. Geophys. Res.-Atmos.*, 99, 22 831–22 855, 1994.
- 10 Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO<sub>2</sub> retrieval algorithm-Part II: Global XCO<sub>2</sub> data characterization, *Atmos. Meas. Tech.*, 5, 687–707, <https://doi.org/10.5194/amt-5-687-2012>, <http://www.atmos-meas-tech.net/5/687/2012/>, 2012.
- 15 Dai, A.: Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008, *J. Geophys. Res.-Atmos.*, 116, n/a–n/a, <https://doi.org/10.1029/2010JD015541>, <http://dx.doi.org/10.1029/2010JD015541>, d12115, 2011.
- Dai, A.: Dai Global Palmer Drought Severity Index (PDSI). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory., <https://doi.org/10.5065/D6QF8R93>, accessed 15 Aug 2017, 2017.
- Dai, A., Trenberth, K. E., and Qian, T.: A global dataset of Palmer Drought Severity Index for 1870–2002: relationship with soil moisture and effects of surface warming, *Journal of Hydrometeorology*, 5, 1117–1130, 2004.
- Damm, A., Guanter, L., Paul-Limoges, E., Van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C., and Schaeppman, M.: Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches, *Remote Sens. Environ.*, 166, 91–105, 2015.
- Deng, F., Jones, D., Henze, D., Bousserez, N., Bowman, K., Fisher, J., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P., et al.: Inferring regional sources and sinks of atmospheric CO<sub>2</sub> from GOSAT XCO<sub>2</sub> data, *Atmos. Chem. Phys.*, 14, 3703–3727, 2014.
- 25 Deng, F., Jones, D. B. A., Walker, T. W., Keller, M., Bowman, K. W., Henze, D. K., Nassar, R., Kort, E. A., Wofsy, S. C., Walker, K. A., Bourassa, A. E., and Degenstein, D. A.: Sensitivity analysis of the potential impact of discrepancies in stratosphere–troposphere exchange on inferred sources and sinks of CO<sub>2</sub>, *Atmospheric Chemistry and Physics*, 15, 11 773–11 788, <https://doi.org/10.5194/acp-15-11773-2015>, <https://www.atmos-chem-phys.net/15/11773/2015/>, 2015.
- 30 Deng, F., Jones, D., O'Dell, C. W., Nassar, R., and Parazoo, N. C.: Combining GOSAT XCO<sub>2</sub> observations over land and ocean to improve regional CO<sub>2</sub> flux estimates, *J. Geophys. Res.-Atmos.*, 121, 1896–1913, <https://doi.org/10.1002/2015JD024157>, <http://dx.doi.org/10.1002/2015JD024157>, 2016.
- Fang, Y., Michalak, A. M., Schwalm, C. R., Huntzinger, D. N., Berry, J. A., Ciais, P., Piao, S., Poulter, B., Fisher, J. B., Cook, R. B., Hayes, D., Huang, M., Ito, A., Jain, A., Lei, H., Lu, C., Mao, J., Parazoo, N. C., Peng, S., Ricciuto, D. M., Shi, X., Tao, B., Tian, H., Wang, W., Wei, Y., and Yang, J.: Global land carbon sink response to temperature and precipitation varies with ENSO phase, *Environmental Research Letters*, 12, 064 007, <https://doi.org/10.1088/1748-9326/aa6e8e>, 2017.

- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Van der Velde, M., Vicca, S., Babst, F., et al.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, *Global Change Biology*, 21, 2861–2880, 2015.
- Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., et al.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, *Geophys. Res. Lett.*, 38, 2011.
- Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., Suto, H., and Grainger, R. G.: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements, *Remote Sens. Environ.*, 121, 236–251, <https://doi.org/10.1016/j.rse.2012.02.006>, 2012.
- 10 Guerlet, S., Basu, S., Butz, A., Krol, M., Hahne, P., Houweling, S., Hasekamp, O., and Aben, I.: Reduced carbon uptake during the 2010 Northern Hemisphere summer from GOSAT, *Geophys. Res. Lett.*, 40, 2378–2383, 2013.
- Gutschick, V. P. and BassiriRad, H.: Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences, *New Phytologist*, 160, 21–42, 2003.
- Harper, A. B., Wiltshire, A. J., Cox, P. M., Friedlingstein, P., Jones, C. D., Mercado, L. M., Sitch, S., Williams, K., and Duran-Rojas, C.: Vegetation distribution and terrestrial carbon cycle in a carbon cycle configuration of JULES4.6 with new plant functional types, *Geosci. Model Dev.*, 11, 2857–2873, <https://doi.org/10.5194/gmd-11-2857-2018>, <https://www.geosci-model-dev.net/11/2857/2018/>, 2018.
- 15 He, Z., Lei, L., Welp, L. R., Zeng, Z.-C., Bie, N., Yang, S., and Liu, L.: Detection of Spatiotemporal Extreme Changes in Atmospheric CO<sub>2</sub> Concentration Based on Satellite Observations, *Remote Sensing*, 10, 839, 2018.
- Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, 7, 2413–2433, 2007.
- 20 Huntzinger, D., Michalak, A., Schwalm, C., Ciais, P., King, A., Fang, Y., Schaefer, K., Wei, Y., Cook, R., Fisher, J., et al.: Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions, *Scientific Reports*, 7, 2017.
- Ishizawa, M., Mabuchi, K., Shirai, T., Inoue, M., Morino, I., Uchino, O., Yoshida, Y., Belikov, D., and Maksyutov, S.: Inter-annual variability of summertime CO<sub>2</sub> exchange in Northern Eurasia inferred from GOSAT XCO<sub>2</sub>, *Environ. Res. Lett.*, 11, 105001, 2016.
- Ito, A.: Changing ecophysiological processes and carbon budget in East Asian ecosystems under near-future changes in climate: implications for long-term monitoring from a process-based model, *J. Plant Res.*, 123, 577–588, 2010.
- 25 Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A., Middleton, E., Huemmrich, K., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2, *Atmos. Meas. Tech.*, 6, 2803–2823, <https://doi.org/10.5194/amt-6-2803-2013>, 2013.
- Joiner, J., Yoshida, Y., Guanter, L., and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: simulations and application to GOME-2 and SCIAMACHY, *Atmos. Meas. Tech.*, 9, 3939–3967, <https://doi.org/10.5194/amt-9-3939-2016>, <https://www.atmos-meas-tech.net/9/3939/2016/>, 2016.
- 30 Jones, C. D. and Cox, P. M.: On the significance of atmospheric CO<sub>2</sub> growth rate anomalies in 2002–2003, *Geophys. Res. Lett.*, 32, n/a–n/a, <https://doi.org/10.1029/2005GL023027>, 114816, 2005.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., et al.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.-Biogeo.*, 116, 2011.
- 35 Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., et al.: Compensatory water effects link yearly global land CO<sub>2</sub> sink changes to temperature, *Nature*, 541, 516–520, 2017.

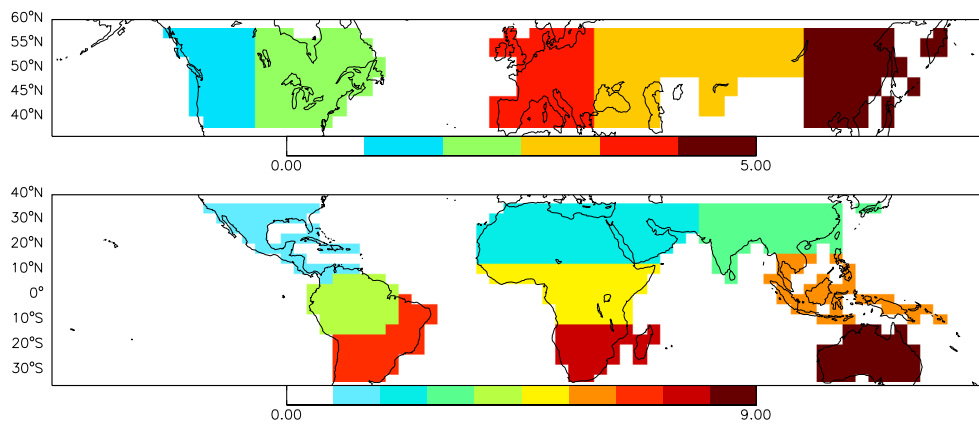
- Keeling, C. D., Adams Jr, J. A., Ekdahl Jr, C. A., and Guenther, P. R.: Atmospheric carbon dioxide variations at the South Pole, *Tellus*, 28, 552–564, 1976a.
- Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Ekdahl Jr, C. A., Guenther, P. R., Waterman, L. S., and Chin, J. F.: Atmospheric carbon dioxide variations at Mauna Loa observatory, Hawaii, *Tellus*, 28, 538–551, 1976b.
- 5 Keeling, C. D., Whorf, T. P., Wahlen, M., and Van der Plichtt, J.: Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980, *Nature*, 375, 666–670, 1995.
- Keppel-Aleks, G., Wolf, A. S., Mu, M., Doney, S. C., Morton, D. C., Kasibhatla, P. S., Miller, J. B., Dlugokencky, E. J., and Randerson, J. T.: Separating the influence of temperature, drought, and fire on interannual variability in atmospheric CO<sub>2</sub>, *Global Biogeochem By.*, 28, 1295–1310, 2014.
- 10 Kondo, M., Ichii, K., Takagi, H., and Sasakawa, M.: Comparison of the data-driven top-down and bottom-up global terrestrial CO<sub>2</sub> exchanges: GOSAT CO<sub>2</sub> inversion and empirical eddy flux upscaling, *J. Geophys. Res.-Biogeo.*, 120, 1226–1245, 2015.
- Krol, M., Houweling, S., Bregman, B., Broek, M., Segers, A., Velthoven, P. v., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, *Atmos. Chem. Phys.*, 5, 417–432, <https://doi.org/10.5194/acp-5-417-2005>, 2005.
- 15 Kulawik, S., Wunch, D., O'Dell, C., Frankenberg, C., Reuter, M., Oda, T., Chevallier, F., Sherlock, V., Buchwitz, M., Osterman, G., Miller, C. E., Wennberg, P. O., Griffith, D., Morino, I., Dubey, M. K., Deutscher, N. M., Notholt, J., Hase, F., Warneke, T., Sussmann, R., Robinson, J., Strong, K., Schneider, M., De Mazière, M., Shiomi, K., Feist, D. G., Iraci, L. T., and Wolf, J.: Consistent evaluation of ACOS-GOSAT, BESD-SCIAMACHY, CarbonTracker, and MACC through comparisons to TCCON, *Atmos. Meas. Tech.*, 9, 683–709, <https://doi.org/10.5194/amt-9-683-2016>, <http://www.atmos-meas-tech.net/9/683/2016/>, 2016.
- 20 Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., et al.: Justify your alpha, *Nat. Hum. Behav.*, p. 1, 2018.
- Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., Collatz, G. J., Menemenlis, D., Ott, L., Pawson, S., et al.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO<sub>2</sub> sampling on the inference of terrestrial biospheric sources and sinks, *Tellus B*, 66, 22 486, <https://doi.org/10.3402/tellusb.v66.22486>, 2014.
- 25 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*, 358, <https://doi.org/10.1126/science.aam5690>, <http://science.sciencemag.org/content/358/6360/eaam5690>, 2017.
- Mabuchi, K., Takagi, H., and Maksyutov, S.: Relationships between CO<sub>2</sub> flux estimated by inverse analysis and land surface elements in South America and Africa, *J. Meteorol. Soc. Jpn*, 94, 415–430, 2016.
- 30 Maksyutov, S., Takagi, H., Valsala, V. K., Saito, M., Oda, T., Saeki, T., Belikov, D. A., Saito, R., Ito, A., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., and Yokota, T.: Regional CO<sub>2</sub> flux estimates for 2009–2010 based on GOSAT and ground-based CO<sub>2</sub> observations, *Atmos. Chem. Phys.*, 13, 9351–9373, <https://doi.org/10.5194/acp-13-9351-2013>, <https://www.atmos-chem-phys.net/13/9351/2013/>, 2013.
- Marcolla, B., Rödenbeck, C., and Cescatti, A.: Patterns and controls of inter-annual variability in the terrestrial carbon budget, *Biogeosciences*, 14, 3815–3829, <https://doi.org/10.5194/bg-14-3815-2017>, <https://www.biogeosciences.net/14/3815/2017/>, 2017.
- 35 Monsi, M. and Saeki, T.: Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduktion, *Jpn. J. Bot.*, 14, 22–52, 1953.
- NASA-SIF: NASA GOME-2 Solar Induced Fluorescence product, <http://avdc.gsfc.nasa.gov/>, accessed: 2016-07-25, 2016.

- Nassar, R., Jones, D., Kulawik, S., Worden, J., Bowman, K., Andres, R. J., Suntharalingam, P., Chen, J., Brenninkmeijer, C., Schuck, T., et al.: Inverse modeling of CO<sub>2</sub> sources and sinks using satellite observations of CO<sub>2</sub> from TES and surface flask measurements, *Atmos. Chem. Phys.*, 11, 6029–6047, 2011.
- O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO<sub>2</sub> retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, 5, 99–121, <https://doi.org/10.5194/amt-5-99-2012>, <http://www.atmos-meas-tech.net/5/99/2012/>, 2012.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., et al.: The JRA-25 reanalysis, *Journal of the Meteorological Society of Japan. Ser. II*, 85, 369–432, 2007.
- 10 Papageorgiou, G. C. and Govindjee: Chlorophyll a fluorescence: a signature of photosynthesis, vol. 19, Springer Science & Business Media, 2007.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Pétron, G., Hirsch, A. I., et al.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci.*, 104, 18925–18930, <https://doi.org/10.1073/pnas.0708986104>, 2007.
- 15 Peylin, P., Law, R., Gurney, K., Chevallier, F., Jacobson, A., Maki, T., Niwa, Y., Patra, P., Peters, W., Rayner, P., et al.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO<sub>2</sub> inversions, *Biogeosciences*, 10, 6699–6720, 2013.
- Polavarapu, S. M., Deng, F., Byrne, B., Jones, D. B. A., and Neish, M.: A comparison of atmospheric CO<sub>2</sub> flux signals obtained from GEOS-Chem flux inversions constrained by in situ or GOSAT observations, *Atmos. Chem. Phys. Discussions*, 2018, 1–55, <https://doi.org/10.5194/acp-2017-1235>, <https://www.atmos-chem-phys-discuss.net/acp-2017-1235/>, 2018.
- 20 Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: a process model based on global satellite and surface data, *Global Biogeochem. Cy.*, 7, 811–841, <https://doi.org/10.1029/93GB02725>, 1993.
- Randerson, J. T., Thompson, M. V., Malmstrom, C. M., Field, C. B., and Fung, I. Y.: Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO<sub>2</sub>, *Global Biogeochem. Cy.*, 10, 585–602, <https://doi.org/10.1029/96GB01981>,
- 25 1996.
- Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, *Journal of Climate*, 24, 6322–6338, 2011.
- Reichle, R. H., Draper, C. S., Liu, Q., Giroto, M., Mahanama, S. P., Koster, R. D., and De Lannoy, G. J.: Assessment of MERRA-2 land surface hydrology estimates, *Journal of Climate*, 30, 2937–2960, 2017.
- 30 Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO<sub>2</sub> flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, 3, 1919–1964, <https://doi.org/10.5194/acp-3-1919-2003>, <https://www.atmos-chem-phys.net/3/1919/2003/>, 2003.
- Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to inter-annual climatic variations? A quantification based on atmospheric CO<sub>2</sub> data, *Biogeosciences*, 15, 2481–2498, <https://doi.org/10.5194/bg-15-2481-2018>,
- 35 <https://www.biogeosciences.net/15/2481/2018/>, 2018.
- Saito, M., Ito, A., and Maksyutov, S.: Optimization of a prognostic biosphere model for terrestrial biomass and atmospheric CO<sub>2</sub> variability, *Geosci. Model Dev.*, 7, 1829–1840, 2014.

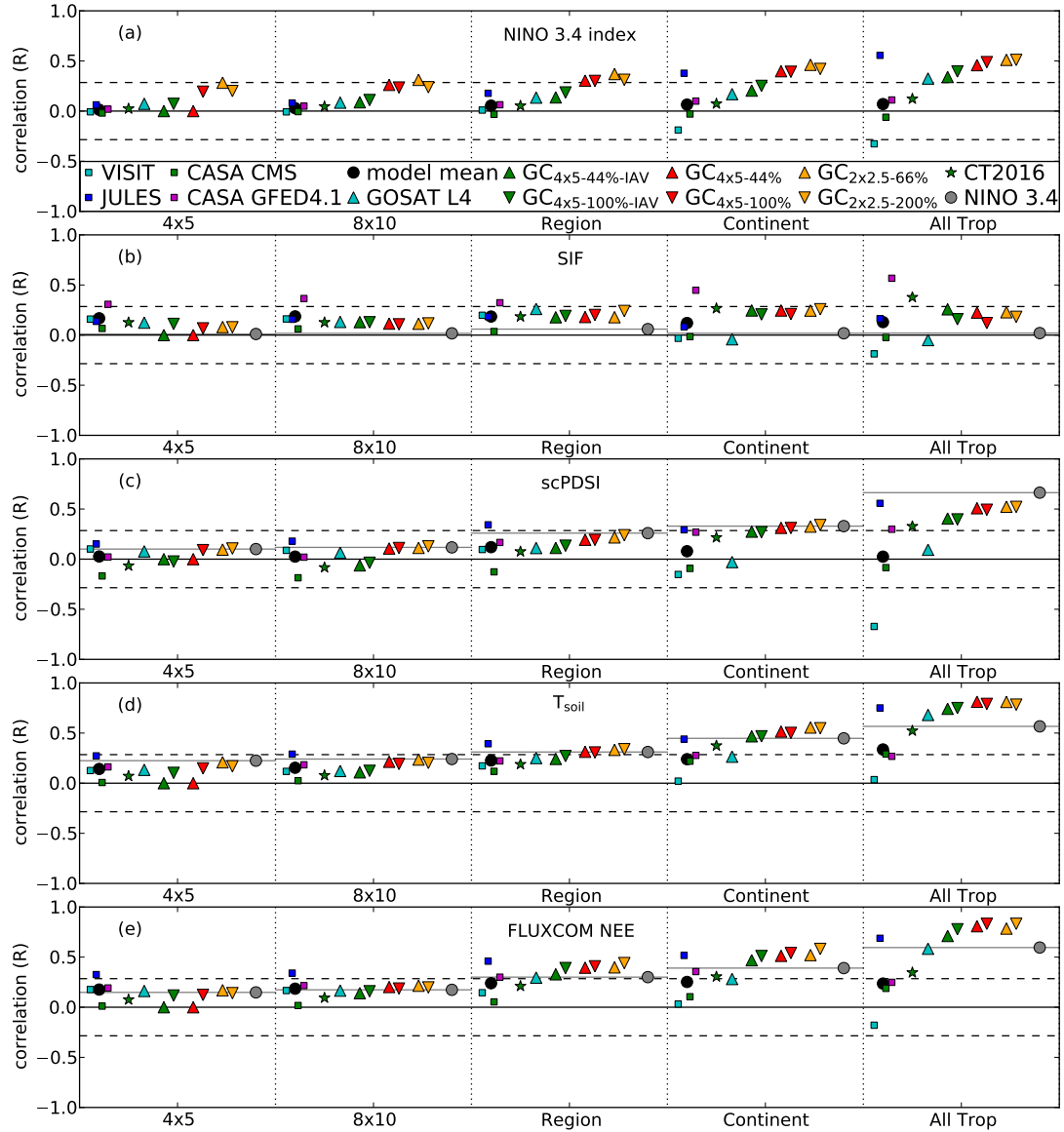
- Schwalm, C. R., Anderegg, W. R., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., et al.: Global patterns of drought recovery, *Nature*, 548, 202, 2017.
- Shiga, Y. P., Michalak, A. M., Fang, Y., Schaefer, K., Andrews, A. E., Huntzinger, D. H., Schwalm, C. R., Thoning, K., and Wei, Y.: Forests dominate the interannual variability of the North American carbon sink, *Environmental Research Letters*, 13, 084015, <https://doi.org/10.1088/1748-9326/aad505>, <http://stacks.iop.org/1748-9326/13/i=8/a=084015>, 2018a.
- Shiga, Y. P., Tadić, J. M., Qiu, X., Yadav, V., Andrews, A. E., Berry, J. A., and Michalak, A. M.: Atmospheric CO<sub>2</sub> Observations Reveal Strong Correlation Between Regional Net Biospheric Carbon Uptake and Solar-Induced Chlorophyll Fluorescence, *Geophys. Res. Lett.*, 45, 1122–1132, <https://doi.org/10.1002/2017GL076630>, 2018b.
- Sippel, S., Reichstein, M., Ma, X., Mahecha, M. D., Lange, H., Flach, M., and Frank, D.: Drought, Heat, and the Carbon Cycle: a Review, *Current Climate Change Reports*, pp. 1–21, 2018.
- Smith, M. D.: An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research, *Journal of Ecology*, 99, 656–663, 2011.
- Stanevich, I.: Characterizing model errors in chemical transport modelling of methane: Impact of coarse model resolution, 2018.
- Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., et al.: Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea–air CO<sub>2</sub> flux over the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, 56, 554–577, 2009.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences*, 13, 4291–4313, <https://doi.org/10.5194/bg-13-4291-2016>, <https://www.biogeosciences.net/13/4291/2016/>, 2016.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr, A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, 6, 3423–3441, 2006.
- van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, *Earth System Science Data*, 9, 697–720, <https://doi.org/10.5194/essd-9-697-2017>, <https://www.earth-syst-sci-data.net/9/697/2017/>, 2017.
- Wang, W., Ciais, P., Nemani, R. R., Canadell, J. G., Piao, S., Sitch, S., White, M. A., Hashimoto, H., Milesi, C., and Myneni, R. B.: Variations in atmospheric CO<sub>2</sub> growth rates coupled with tropical temperature, *P. Natl. Acad. Sci.*, 110, 13 061–13 066, 2013.
- Wood, J. D., Griffis, T. J., Baker, J. M., Frankenberg, C., Verma, M., and Yuen, K.: Multiscale analyses of solar-induced fluorescence and gross primary production, *Geophys. Res. Lett.*, 44, 533–541, <https://doi.org/10.1002/2016GL070775>, <http://dx.doi.org/10.1002/2016GL070775>, 2016GL070775, 2017.
- Wunch, D., Wennberg, P., Messerschmidt, J., Parazoo, N., Toon, G., Deutscher, N., Keppel-Aleks, G., Roehl, C., Randerson, J., Warneke, T., et al.: The covariation of Northern Hemisphere summertime CO<sub>2</sub> with surface temperature in boreal regions, *Atmos. Chem. Phys.*, 13, 9447–9459, 2013.
- Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., Munger, J. W., Kornfeld, A., and Richardson, A. D.: Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest, *Geophys. Res. Lett.*, 42, 2977–2987, 2015.

- Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., and Long, M. S.: Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology, *Geosci. Model Dev.*, 11, 305–319, <https://doi.org/10.5194/gmd-11-305-2018>, <https://www.geosci-model-dev.net/11/305/2018/>, 2018.
- Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., and Zhang, F.: Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications, *Remote Sens. Environ.*, 187, 145–155, 2016a.
- Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., Joiner, J., Guanter, L., Zhang, Y., Zhang, G., et al.: Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America, *Remote Sens. Environ.*, 183, 154–169, 2016b.
- 10 Zhang, Y., Joiner, J., Gentine, P., and Zhou, S.: Reduced solar-induced chlorophyll fluorescence from GOME-2 during Amazon drought caused by dataset artifacts, *Global change biology*, 24, 2229–2230, 2018.
- Zscheischler, J., Mahecha, M. D., Von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J. T., Schölkopf, B., Seneviratne, S. I., Tomelleri, E., et al.: A few extreme events dominate global interannual variability in gross primary production, *Environ. Res. Lett.*, 9, 035001, 2014.





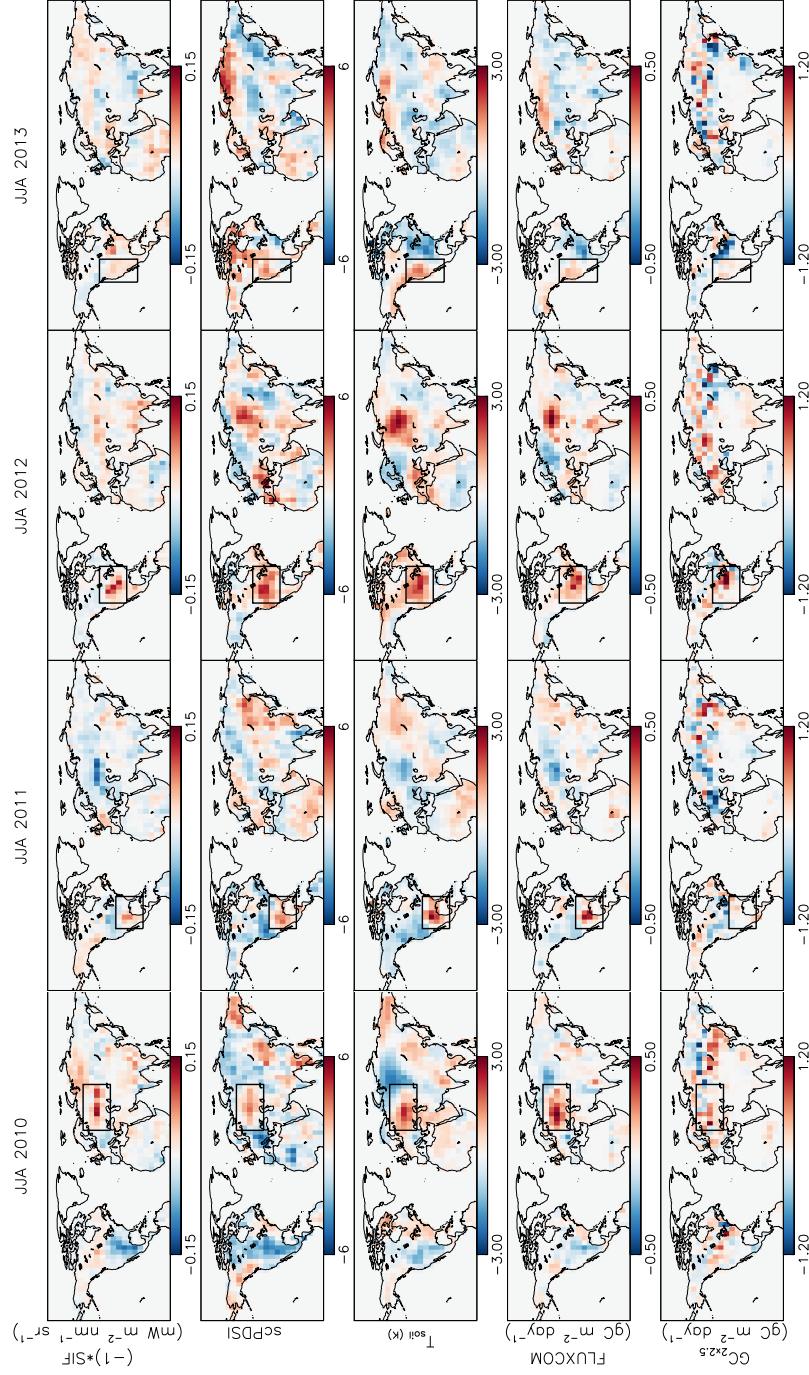
**Figure 1.** Land area at  $4^\circ \times 5^\circ$  resolution grouped into sub-continental regions in (a) the northern extratropics and (b) the tropics. In the tropics, we generate three continents by combining the regions in the Americas, Africa and the Middle East, and the Asia-Pacific and Indian sub-continent.



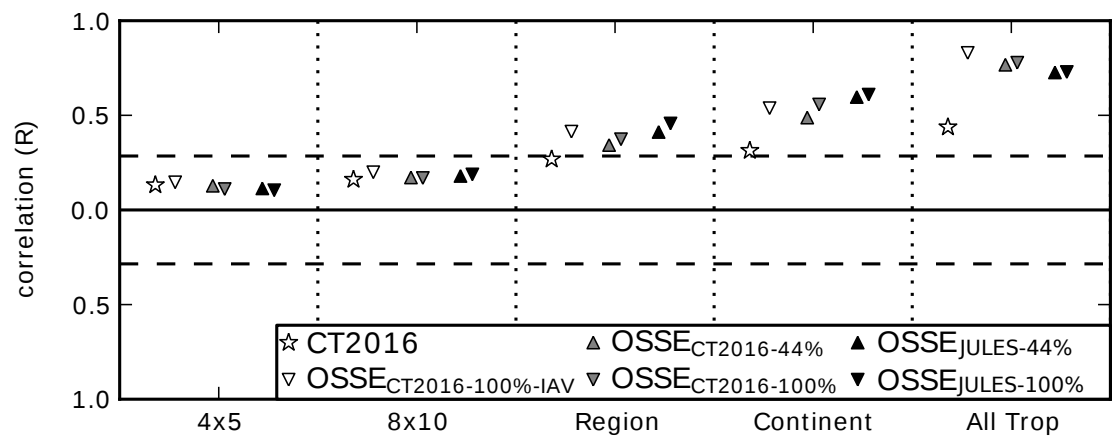
**Figure 2.** Correlation in the tropics over a range of scales for models and inversions with (top) NINO 3.4 index, (second)  $(-1) \times \text{SIF}$ , (third) scPDSI, (fourth)  $T_{\text{soil}}$ , and (bottom) FLUXCOM NEE in the tropics. Squares represent correlations with terrestrial ecosystem model NEE IAV: VISIT (cyan), JULES (blue), CASA GFED CMS (green), CASA GFED 4.1 (magenta) and the black circle shows the mean correlation of the models. Triangles represent correlations with the GOSAT flux inversion NEE IAV: GOSAT L4 (cyan up-triangle),  $\text{GC}_{4 \times 5-44\%-\text{IAV}}$  NEE IAV (green up-triangle),  $\text{GC}_{4 \times 5-100\%-\text{IAV}}$  NEE IAV (green down-triangle),  $\text{GC}_{4 \times 5-44\%}$  NEE IAV (red up-triangle),  $\text{GC}_{4 \times 5-100\%}$  NEE IAV (red down-triangle),  $\text{GC}_{2 \times 2.5-66\%}$  NEE IAV (orange up-triangle), and  $\text{GC}_{2 \times 2.5-200\%}$  NEE IAV (orange down-triangle). The green star show the correlation with CT2016 NEE IAV. The grey circle shows the correlation with the NINO 3.4 index. Dashed black lines indicate the correlation required for an  $\alpha$  of 0.05, therefore, all correlations greater than the dashed black line indicate  $P < 0.05$ .

Correlations of monthly anomalies over tropical land at  $4^\circ \times 5^\circ$  spatial resolution. Columns show coefficient of correlation ( $R$ ) of (left)  $\text{GC}_{2 \times 2.5-200\%}$  NEE IAV, (center) NINO 3.4 index, and (right) the difference between the two with (top row) the NINO 3.4 index, (second row)  $(-1) \times \text{SIF}$ , (third row)  $\text{sePDSI}$ , (fourth row)  $T_{\text{soil}}$ , and (bottom row) FLUXCOM NEE.

5 NEE anomalies ( $\text{gC m}^{-2} \text{ day}^{-1}$ ) for FLUXCOM and  $\text{GC}_{2 \times 2.5-200\%}$  in the tropics. (left column) Monthly anomalies, (center column) smoothed (3-month running mean) monthly anomalies, and (right column)  $\text{DIFF}_{\text{continent=tropics}}$  (see Sect. ?? to see how this is calculated) for (a–b) the entire tropics, (c–e) the Americas, (f–h) Africa and the Middle East, and (i–k) the Asia Pacific and Indian sub-continent. For each sub-plot,  $R^2$  shows the coefficient of determination between  $\text{GC}_{2 \times 2.5-200\%}$  and FLUXCOM NEE anomalies within the sub-plot.



**Figure 3.** Northern extratropical anomalies during JJA. Anomalies for (top row)  $(-1) \times \text{SIF}$ , (second)  $\text{scPDSI}$ , (third)  $T_{\text{soil}}$ , (fourth)  $\text{FLUXCOM NEE}$ , and (bottom)  $\text{GC}_{2x2.5-200\% \text{ NEE}}$  over JJA for (left to right) 2010–2013. Black boxes highlight major climate anomalies: the 2010 Russian heat wave, 2011 drought in Mexico and southern USA, the 2012 North American drought, and the 2013 California drought.



**Figure 4.** Mean correlation coefficient ( $R$ ) with the true OSSE NEE IAV over a range of spatial scales for CT2016 NEE IAV (white star), OSSE<sub>CT2016-100%-IAV</sub> NEE IAV (white down-triangle), OSSE<sub>CT2016-44%</sub> NEE IAV (grey up-triangle), OSSE<sub>CT2016-100%</sub> NEE IAV (grey down-triangle), OSSE<sub>JULES-44%</sub> NEE IAV (black up-triangle), and OSSE<sub>JULES-100%</sub> (black down-triangle) NEE IAV.

Monthly NEE anomalies ( $\text{gC m}^{-2} \text{ day}^{-1}$ ) for OSSE<sub>JULES=100%</sub> (red), OSSE<sub>CT2016=100%</sub> (green), OSSE<sub>CT2016=100%=IAV</sub> (blue) and true NEE IAV (black) in the tropics. (left column) Monthly anomalies, (center column) smoothed (3-month running mean) monthly anomalies, and (right column)  $\text{DIFF}_{\text{continent=tropics}}$  (see Sect. ?? to see how this is calculated) for (a–b) the entire tropics, (c–e) the Americas, (f–h) Africa and the Middle East, and (i–k) the Asia Pacific and Indian sub-continent.

5 Correlation between FLUXCOM MARS GPP anomalies and SHF anomalies at  $2^\circ \times 2.5^\circ$  spatial resolution.

Comparison of GC<sub>IAV</sub> posterior and prior NEE IAV. (a) Correlation coefficient ( $R$ ) between the posterior and prior NEE IAV in the tropics at the spatial scale of  $4^\circ \times 5^\circ$ . (b) Mean correlation coefficient ( $R$ ) between posterior and prior NEE IAV in the tropics for different degrees of spatial aggregation. (c) Northern extratropical anomalies during JJA for (top) prior and (bottom) posterior NEE for (left–right columns) 2010–2013.

**Table 1.** Setup of GEOS-Chem flux inversions. Differences are in model transport resolution, prior fluxes, and prior errors.

Name	Resolution	Prior flux error	Prior flux IAV
GC <sub>2×2.5-200%</sub>	2° × 2.5°	200%	No (mean 2010–2013)
GC <sub>2×2.5-66%</sub>	2° × 2.5°	66%	No (mean 2010–2013)
GC <sub>4×5-100%</sub>	4° × 5°	100%	No (mean 2010–2013)
GC <sub>4×5-44%</sub>	4° × 5°	44%	No (mean 2010–2013)
GC <sub>4×5-100%-IAV</sub>	4° × 5°	100%	Yes
GC <sub>4×5-44%-IAV</sub>	4° × 5°	44%	Yes

**Table 2.** Slope and coefficient of determination ( $R^2$ ) for linear regressions of inversion/TBM NEE anomalies against proxy/FLUXCOM anomalies across the entire tropics.

Model/Inversion gC m <sup>-2</sup> day <sup>-1</sup>	FLUXCOM NEE gC m <sup>-2</sup> day <sup>-1</sup>		SIF mW m <sup>-2</sup> nm <sup>-1</sup> sr <sup>-1</sup>		scPDSI		$T_{soil}$ (K)		NINO 3.4 index (K)	
	slope	$R^2$	slope	$R^2$	slope	$R^2$	slope	$R^2$	slope	$R^2$
GC <sub>2×2.5-200%</sub>	1.87	0.69	0.90	0.03	0.078	0.27	0.100	0.61	0.026	0.26
GC <sub>2×2.5-66%</sub>	1.03	0.62	0.65	0.05	0.045	0.27	0.061	0.66	0.015	0.26
GC <sub>4×5-100%</sub>	1.70	0.69	0.54	0.01	0.067	0.24	0.093	0.63	0.022	0.24
GC <sub>4×5-44%</sub>	1.06	0.65	0.65	0.05	0.044	0.26	0.061	0.66	0.014	0.21
GC <sub>4×5-100%-IAV</sub>	2.10	0.61	0.94	0.03	0.071	0.16	0.12	0.56	0.024	0.16
GC <sub>4×5-44%-IAV</sub>	1.57	0.51	0.03	0.00	0.06	0.16	0.087	0.55	0.017	0.12
GOSAT L4	1.59	0.34	-0.30	0.00	0.017	0.01	0.106	0.46	0.020	0.11
GOSAT L4 <sub>w/BB</sub>	1.69	0.33	-0.02	0.00	0.007	0.00	0.107	0.40	0.016	0.06
CT2016	0.66	0.12	1.58	0.14	0.042	0.11	0.057	0.27	0.001	0.02
CT2016 <sub>w/BB</sub>	0.79	0.14	1.73	0.14	0.027	0.04	0.059	0.24	0.001	0.00
VISIT	-0.50	0.03	-1.15	0.04	-0.13	0.45	0.006	0.00	-0.021	0.11
CASA 4.1	0.38	0.06	1.88	0.32	0.030	0.09	0.023	0.07	0.004	0.01
CASA CMS	0.33	0.04	-0.09	0.00	-0.010	0.01	0.029	0.08	-0.002	0.00
JULES	1.85	0.47	0.96	0.027	0.10	0.31	0.116	0.56	0.033	0.31



**Table 3.** Slope and coefficient of determination ( $R^2$ ) for linear regressions of regional-sub-continental region inversion/TBM NEE anomalies against proxy/FLUXCOM anomalies during JJA in the northern extratropics.

Model/Inversion gC m <sup>-2</sup> day <sup>-1</sup>	FLUXCOM NEE gC m <sup>-2</sup> day <sup>-1</sup>		SIF mW m <sup>-2</sup> nm <sup>-1</sup> sr <sup>-1</sup>		scPDSI		$T_{soil}$ (K)	
	slope	$R^2$	slope	$R^2$	slope	$R^2$	slope	$R^2$
GC <sub>2×2.5-200%</sub>	1.56	0.54	4.07	0.14	0.052	0.21	0.17	0.56
GC <sub>2×2.5-66%</sub>	1.28	0.65	3.32	0.16	0.041	0.24	0.13	0.57
GC <sub>4×5-100%</sub>	1.36	0.49	4.13	0.17	0.054	0.28	0.16	0.62
GC <sub>4×5-44%</sub>	1.29	0.64	3.36	0.17	0.045	0.29	0.14	0.65
GC <sub>4×5-100%-IAV</sub>	1.28	0.26	6.8	0.27	0.05	0.16	0.16	0.36
GC <sub>4×5-44%-IAV</sub>	0.79	0.15	4.66	0.20	0.026	0.06	0.10	0.21
GOSAT L4	1.59	0.33	5.86	0.17	0.086	0.35	0.19	0.43
GOSAT L4 <sub>w/BB</sub>	1.59	0.34	6.52	0.21	0.090	0.39	0.18	0.39
CT2016	0.21	0.01	4.03	0.13	0.000	0.00	0.04	0.03
CT2016 <sub>w/BB</sub>	0.18	0.006	4.59	0.16	0.002	0.00	0.03	0.01
VISIT	0.93	0.47	3.25	0.21	0.059	0.67	0.10	0.50
CASA 4.1	0.37	0.12	3.96	0.48	0.020	0.11	0.05	0.20
CASA CMS	0.16	0.01	4.13	0.34	0.00	0.00	0.02	0.02
JULES	1.58	0.29	7.26	0.23	0.075	0.23	0.23	0.52

**Table 4.** Slope and coefficient of determination ( $R^2$ ) for linear regressions of OSSE posterior NEE anomalies against the true NEE IAV and  $\text{OSSE}_{\text{JULES}-100\%}$ .

Tropics				
Inversion	true NEE IAV		$\text{OSSE}_{\text{JULES}-100\%}$	
	slope	$R^2$	slope	$R^2$
$\text{OSSE}_{\text{JULES}-100\%}$	0.67	0.53		
$\text{OSSE}_{\text{JULES}-44\%}$	0.58	0.53	0.91	0.91
$\text{OSSE}_{\text{CT2016}-100\%}$	0.55	0.61	0.84	0.84
$\text{OSSE}_{\text{CT2016}-44\%}$	0.42	0.59	0.69	0.77
$\text{OSSE}_{\text{CT2016}-100\%-\text{IAV}}$	0.75	0.69	0.70	0.48
CT2016	0.31	0.19	0.50	0.15
Northern Extratropics				
Inversion	true NEE IAV		$\text{OSSE}_{\text{JULES}-100\%}$	
	slope	$R^2$	slope	$R^2$
$\text{OSSE}_{\text{JULES}-100\%}$	0.35	0.39		
$\text{OSSE}_{\text{JULES}-44\%}$	0.27	0.48	0.76	0.80
$\text{OSSE}_{\text{CT2016}-100\%}$	0.30	0.30	1.04	0.88
$\text{OSSE}_{\text{CT2016}-44\%}$	0.31	0.43	1.06	0.62
$\text{OSSE}_{\text{CT2016}-100\%-\text{IAV}}$	0.63	0.15	0.55	0.41
CT2016	0.48	0.46	0.18	0.05