Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2017-813-AC1, 2017

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite" *by* Scot M. Miller et al.

Scot M. Miller et al.

scot.m.miller@gmail.com

Received and published: 10 November 2017

We would like to thank the reviewer for ideas and suggestions for the manuscript. This feedback will be very helpful for updating and improving the manuscript. Below, we have included both the reviewer's suggestions (in bold) along with the associated changes we plan to make.

- **I found it particularly difficult to follow the logic of the paper and to evaluate the soundness of the approach. As a preliminary step for publication, the authors should seriously invest in making their study accessible to the**

Printer-friendly version

Discussion paper

broad audience of ACP.

This is very helpful feedback for framing the paper and describing the components of the methodology. We plan to re-frame the paper logic in several ways to make it more accessible to a broad audience. First, we plan to expand the overall description of the paper and description of the general approach at the end of the introduction (pg. 2, line 28 to pg. 3, line 21). We will describe the paper narrative in non-technical terms to give the reader an intuitive, high-level overview of the paper logic and flow. This description would provide better intuition for a wide audience of readers, especially those readers who may skip over the more technical information in the methodology (sect. 2).

Second, we will simplify the methods section (sect. 2) so that it is accessible to a broad audience. For example, this section contains seven equations. We will move several of these equations to the SI (e.g., Eq. 4-7) and instead expand the non-technical portions of the description. In this way, the paper will still include all of the technical detail for readers who want it, but the description in the main paper will be accessible to a broader audience.

Third, we will provide more references to existing studies that use similar approaches. Readers who are interested in more details on the methodology could gain greater context using these references. We will make this change throughout the manuscript and particularly from pg. 6, line 10 to pg. 7, line 20.

- **The paper concludes to a limited utility of OCO-2 retrievals for flux estimation with current retrieval algorithms and transport model. This may be correct, but is orthogonal to the claim made by Liu et al. (2017). The disagreement should be clearly stated.**

  We will discuss this difference in the revised version of the manuscript. Liu et al. (2017) was published after this ACPD manuscript, so it is only now possible to make this comparison. Liu et al. use an atmospheric inversion to estimate

CO_2 fluxes for different tropical regions of the globe. They estimate uncertainties in their regional budget estimates, and those uncertainties are generally smaller than implied by the current ACPD manuscript.

Liu et al. (2017) use a 4DVAR approach and estimate the posterior uncertainties using a small number of Monte Carlo simulations. However, these uncertainty estimates are likely to be underestimates – due to compromises required to make the inversion computationally tractable. For example, most satellite-based inversions like Liu et al. (2017) do not fully account for error correlations or biases in the observations and atmospheric model; these studies typically use a diagonal error covariance matrix. Furthermore, Liu et al. (2017) and other studies use a small number of Monte Carlo simulations to estimate the errors (e.g., 60 simulations in Liu et al. (2014)). By contrast, Ribgy et al. (2011) and Ganesan et al. (2014) argue that 100,000 and 25,000 realizations are necessary to robustly estimate uncertainties for their particular inverse modeling problems. Note that it is not always possible to generate large numbers of realizations or fully account for error correlations in current satellite-based inverse models due to computational constraints. In the ACPD manuscript, we do not use a 4DVAR inverse model for this reason.

Consistent with this interpretation, results from the OCO-2 flux team ongoing intercomparison study indicate much larger uncertainties in estimated fluxes. The results are broadly consistent with those presented in the current ACPD manuscript (e.g., Crowell et al. 2017); preliminary results indicate that OCO-2 observations currently provide robust constraints for hemispheric regions but provide weaker constraints for individual continents or subcontinents. More specifically, recent flux team comparisons include CO_2 flux estimates from about eight different inverse modeling groups, and the level of disagreement among these estimates provides a measure of uncertainty in current top-down flux estimates that use the same version of the OCO-2 retrievals as applied in the current work

and in Liu et al. (2017). These estimates (using nadir observations) often show relatively good agreement for total hemispheric terrestrial CO_2 budgets, with the disagreement among inverse modeling estimates being smaller than the total CO_2 budget for a given hemisphere. The opposite is often true of CO_2 budgets estimated for smaller regions (e.g., Sub-Saharan Africa or Tropical Asia), with the disagreement among inverse modeling estimates usually being larger than the total budget. This ongoing work is consistent with the interpretation in the current manuscript.

- **Section 3.1 and the first part of Section 3.3 reinvent the wheel. See, e.g., Olsen and Randerson (2004) and Worden et al. (2017). Similarly, l. 23-28 are just an adaptation of an old argument (Rayner and O'Brien, 2001).**

  The studies mentioned above investigate several requirements for constraining carbon budgets with satellite observations. Rayner and O'Brien (2001) explore the measurement precision required for space-based constraints on surface CO_2 fluxes. Olsen and Randerson (2004) model XCO_2 column enhancements across the globe due to surface CO_2 fluxes and compare them with surface enhancements. Lastly, Worden et al. (2017) estimate the errors in OCO-2 XCO_2 observations.

  As the reviewer points out, the concepts used in Sects. 3.1 and 3.3 are, in part, built on these earlier approaches. However, the purpose of this section is not to develop new concepts. Rather, we build on existing concepts to assess real OCO-2 data. Rayner and O'Brien (2001) and Olsen and Randerson (2004), by contrast, did not have any real XCO_2 observations at their disposal, only simulations of possible future observations. Furthermore, we feel that these sections provide useful context and improve the manuscript narrative. Much of the manuscript presents the results of statistical experiments. These experiments use, as inputs, XCO_2 observations from OCO-2 and estimates of atmospheric transport and satellite retrieval errors. Sect. 3.1 provides visualizations of those

inputs.

In the revised manuscript, we will cite the studies listed above, clarify that we use concepts from the studies, and explain that we apply those concepts to real observations from OCO-2. We will also compare the retrieval errors in Sect. 3.1 against those in Worden et al. (2017). Lastly, we will shorten the first part of Sect. 3.3. That section presents the synthetic study results with no errors; these results serve as a baseline for subsequent results that do include simulated errors.

- **The retrieval error simulations of Fig. 3 look overly optimistic in comparison to the validation results of Wunch et al. (2017).**

  Wunch et al. (2017) compare OCO-2 XCO_2 retrievals against XCO_2 observations at TCCON sites (the Total Column Observing Network). They report an average site bias of 0.22 ppm for comparisons between land nadir retrievals and TCCON sites. They also report an average root mean squared error of 1.31 ppm for the land nadir and TCCON comparisons (Table 3 in Wunch et al. 2017).

  The errors in Fig. 3c-3f do appear slightly smaller than the numbers reported above. However, the errors in Fig. 3 are the mean of individual sounding errors in February and July, respectively – meaned within each PCTM grid box for an entire month. Hence, the errors displayed in this plot will be somewhat smaller than the errors on individual soundings (as reported in Wunch et al. 2017). By contrast, Fig. 1 shows the standard deviation of the estimated retrieval errors (instead of the mean as in Fig. 3). These standard deviations are larger than the mean and broadly consistent with the errors estimated by Wunch et al. (2017).

  In the revised manuscript, we will clarify that the errors displayed in Fig. 3 are monthly means. Furthermore, we will compare and contrast the estimated errors with those estimated in Wunch et al. (2017) and in Worden et al. (2017).

- **Section 3.2 looks for flux patterns in XCO_2. Most top-down studies from OCO-2 would use a Bayesian approach where flux-error patterns are looked**

for. This is more challenging because the signal is even smaller (while the paragraph in-between p. 5 and p. 6 suggests that the two approaches are rather equivalent with respect to the measurement information content). One should therefore discuss this limitation and further tone down the conclusions of the paper.

The reviewer makes a great point, and we will add a discussion of this point to Sects. 2.2 (pgs. 5-6) and 3.2. The approach used here searches for flux patterns as they manifest in XCO_2. Phrased differently, the approach examines s as seen through the OCO-2 observations, where s are the fluxes. A Bayesian approach, by contrast, estimates $s - s_p$, where s_p is the prior flux estimate. This residual flux ($s - s_p$) is presumably smaller than the total flux (s). As a result, inversions essentially estimate a smaller flux signal than the flux signal examined in this study.

The reviewer's argument could therefore imply more pessimistic results than presented in the current manuscript – that the CO_2 flux constraint is weaker than reported in the present study. This issue, however, may also be more nuanced. If the prior estimate is poor, the residual flux ($s - s_p$) will be large. These large flux patterns should be relatively easy to detect using XCO_2 observations, but the inversion will need to rely heavily on the XCO_2 observations (and not on the prior) to make a robust posterior estimate. By contrast, if the prior estimate is very accurate, the residual flux ($s - s_p$) will be small. The inversion will need to estimate a small flux signal, a signal that may be difficult to parse using XCO_2 observations. However, the posterior flux estimate will still be relatively robust due to the accurate prior.

Furthermore, this issue is specific to the setup of each individual inverse model. For example, existing CO_2 inversions use a wide variety of prior flux models. Mueller et al. (2008) use a non-informative prior (i.e., a flat prior) that contributes little information on the fluxes. In addition, many geostatistical inverse modeling

studies use environmental driver data in place of a tradition prior flux estimate (e.g., Gourdji et al. 2008; 2012). These studies choose environmental driver data using a model selection approach in a manner that is somewhat akin to the current ACPD manuscript. In the present study, we instead try to examine more fundamental questions about the robustness of the flux constraint, questions that are independent of subjective choices specific to each inverse model setup.

References

Crowell, S., et al. (2017, October). The OCO-2 Level 4 Flux Product. Poster presented at the OCO-2 Science Team Meeting, Boulder, Colorado.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. E., Sturges, W. T., Moncrieff, J. B., O'Doherty, S. (2015). Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network. Atmospheric Chemistry and Physics, 15(11), 6393–6406, doi:10.5194/acp-15-6393-2015.

Gourdji, S. M., Mueller, K. L., Schaefer, K., Michalak, A. M. (2008). Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data. Journal of Geophysical Research: Atmospheres, 113(D21), doi:10.1029/2007JD009733.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., Michalak, A. M. (2012). North American CO_2 exchange: intercomparison of modeled estimates with results from a fine-scale atmospheric inversion. Biogeosciences, 9(1), 457–475, doi:10.5194/bg-9-457-2012.

Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., Collatz, G. J., Menemenlis, D., Ott, L., Pawson, S., Jones, D., Nassar, R. (2014). Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO_2 sampling on the inference of terrestrial biospheric sources and sinks. Tellus B: Chemical and Physical

Meteorology, 66(1), 22486, doi:10.3402/tellusb.v66.22486.

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, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., Eldering, A. (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño. Science, 358(6360), doi:10.1126/science.aam5690.

Mueller, K. L., Gourdji, S. M., Michalak, A. M. (2008). Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements. Journal of Geophysical Research: Atmospheres, 113(D21), doi:10.1029/2007JD009734.

Olsen, S. C. Randerson, J. T. (2004). Differences between surface and column atmospheric CO_2 and implications for carbon cycle research. Journal of Geophysical Research: Atmospheres, 109(D2), doi:10.1029/2003JD003968.

Rayner, P. J. O'Brien, D. M. (2001). The utility of remotely sensed CO_2 concentration data in surface source inversions. Geophysical Research Letters, 28(1), 175–178, doi:10.1029/2000GL011912.

Rigby, M., Manning, A. J., Prinn, R. G. (2011). Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models. Atmospheric Chemistry and Physics, 11(18), 9887–9898, doi:10.5194/acp-11-9887-2011.

Worden, J. R., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., O'Dell, C., Bowman, K. (2017). Evaluation and attribution of OCO-2 XCO_2 uncertainties. Atmospheric Measurement Techniques, 10(7), 2759–2771, doi:10.5194/amt-10-2759-2017.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R.,

Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., Garc, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., Eldering, A. (2017). Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON. Atmospheric Measurement Techniques, 10(6), 2209–2238, doi:10.5194/amt-10-2209-2017.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2017-813, 2017.

Printer-friendly version

Discussion paper

Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2017-813-AC2, 2017

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite" *by* Scot M. Miller et al.

Scot M. Miller et al.

scot.m.miller@gmail.com

Received and published: 13 December 2017

We would like to thank the reviewer for providing ideas and suggestions. These suggestions will be very helpful as we revise the manuscript. Below, we have listed each of the reviewer's comments (in bold) and the associated changes we plan to make to the manuscript.

- **In Figure 3 the single sounding error of OCO-2 is compared to the signal from uncertainties in biospheric CO2 fluxes. The question is if this comparison makes much sense, since the error budget of OCO-2 has a large**

Printer-friendly version

Discussion paper

random component. The impact of biospheric flux uncertainties is more coherent in space and time, i.e. has very different statistics. Because of this the signal/noise ratio could look very different after space-time averaging of the data.

Figure 3 in the current ACPD manuscript shows the mean of all soundings in each PCTM model grid box for February and July, respectively. Reviewer 1 brought up this question as well, and we will clarify this point in the revised manuscript.

As the reviewer points out, the signal-to-noise ratio in Fig. 3 will vary depending on space-time averaging. With that said, many inverse modeling studies report monthly CO_2 flux totals, so the monthly averaging in Fig. 3 is particularly pertinent. Furthermore, the uncertainties in top-down CO_2 flux estimates change when averaged to aggregate space-times scales, so this issue is also a consideration in inverse modeling, not just the analysis in Fig. 3.

We will revise the discussion of Fig. 3 in several ways to account for the reviewer's suggestion. First, we will explain that the signal-to-noise ratio varies depending upon the space and time scales considered, and we will explain why this monthly scale is a particularly useful time period to examine. Second, we will emphasize that this signal-to-noise ratio provides a useful intuition or feel for the data, but we will point out that top-down inverse models leverage the signal in much more sophisticated ways. The limitations of this signal-to-noise comparison thus motivate subsequent analyses in the manuscript.

- **It is not clear to me what fraction of the flux uncertainty space is spanned by the flux patterns that are used in the regression. Probably many of the patterns are not independent, in which case it is not a surprise that many are not selected. This probably goes back to the question whether the range of estimates of the underlying models provides a fair estimate of the overall uncertainty. This is not easy to prove, but with only a single ocean pattern and a single anthropogenic emission pattern it seems conceivable that the**

uncertainty space is underestimated (by the way, how about uncertainties in land-use change?). Some discussion is needed of how such factors may influence the results, and what the implication could be for the estimated OCO-2 performance.

This factor can influence the results, and we will add a discussion of this point to the manuscript. We explore this possibility in the synthetic data experiments (Fig. 5b in the ACPD manuscript). In that experiment, we create synthetic XCO_2 observations using the SiBCASA flux model and an atmospheric transport model. We then run model selection, but we do not include SiBCASA as a possible predictor variable in the regression. In other words, model selection can include several different terrestrial biosphere models (TBMs) in the regression, but it cannot include the TBM that was used to generate the synthetic data in the first place. Fig. 5b in the current ACPD manuscript shows the result. Model selection does not select patterns in every region and every month, but it still selects flux patterns for most regions and months.

This issue also affects Bayesian inverse models. These inversions use a prior flux estimate as an initial guess for the fluxes. If the prior flux estimate is inaccurate, the prior error covariance matrix will have large variances/covariances, and the posterior uncertainties will likely be large. If the prior flux estimate is skilled, the prior error covariance matrix will have small variances/covariances, and the posterior uncertainties will be smaller. In other words, the availability and skill of prior flux models (i.e., TBMs) affects the robustness and uncertainty of the inverse modeling estimate.

- **SPECIFIC COMMENTS**
- **page 1, line 23: 'unlike previous missions' .. but this was the case also for GOSAT and SCIAMACHY.**

We will change the text accordingly. In the revised text, we will remove the

phrase “Unlike previous missions” and briefly explain the similarities and differences among OCO-2, GOSAT, and SCIAMACHY.

- **page 2, line 12: references are needed to the recent special issue on OCO-2 in Science.**

  We will rewrite this paragraph and discuss studies from the new *Science* special issue. This special issue was published after the present ACPD manuscript, and it is now possible to reference these papers in the manuscript.

- **page 3, line 17-20: unless ’region’ is defined more quantitatively these sentences are too vague.**

  We will revise these sentences accordingly. In response to feedback from reviewer 1, we plan to re-write the second half of Sect. 1 to describe the overall objectives and approach in a way that is more accessible to a broad audience. To that end, we will more concisely define the word “region”.

- **page 3, line 19-23: Explain the motivation for this second approach? Is one considered to be more realistic than the other?**

  We will clarify the text in this paragraph. We do not consider one approach to be more accurate than another per se. Rather, it is challenging to estimate realistic retrieval errors because these errors are unknown (except possibly at TCCON sites). We asked several colleagues for advice on how to estimate these errors, and different colleagues recommended different approaches that produce different retrieval error estimates. As a result, we decided to use two different retrieval error estimates – to ensure that the results were not contingent upon the specific method used.

- **page 5, line 9: the constant fluxes need to be defined more quantitatively. What did you use? The same flux for each region and month? Are they**

estimated per region? Does it mean that the regressed flux patterns have zero mean? If so please mention.

We will clarify this topic in the manuscript, and we will define these constant terms more quantitatively.

The constant flux is estimated for each region and each month. This constant flux is included as a predictor variable in the regression, and the regression framework scales the magnitude of the constant flux in each region and month to match the observations.

Equations 1 and 2 in the manuscript describe the overall regression and illustrate these relationships quantitatively:

$$Y = h(\mathbf{X})$$

$$\boldsymbol{z} = \mathbf{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{X}$ are the predictor variables in the regression, $h()$ is the atmospheric transport model, $\boldsymbol{z}$ are the observations, $\boldsymbol{\beta}$ are the coefficients estimated by the regression, and $\boldsymbol{\epsilon}$ are the regression residuals. In this setup, the constant flux terms are individual columns in $\mathbf{X}$. Each column has a value of one in a given region and month and has values of zero elsewhere. Phrased differently, these constant flux terms are analogous to the y-intercept terms in the regression. Also of note, the regression residuals $\boldsymbol{\epsilon}$ have a mean of zero, but the regressed flux patterns will not have a zero mean.

We will make several changes to clarify this topic in the manuscript. We will move Eqs. 1-2 earlier in Sect. 2.2 and describe these equations alongside the description of the constant or intercept terms. In response to reviewer 1, we will move several equations to the SI and simplify the description in Sect. 2.2. Instead, we will dedicate more description to Eqs. 1-2 and will explain how the different predictor variables (including the constant or intercept terms) fit into these equations.

- **page 7, line 25: should we conclude that OCO-2's glint mode retrievals do not provide significant independent information?**

  We will provide additional discussion of this point in the manuscript. The statement above may be too bold to make in the manuscript, especially in context of the reviewer's next suggestion below. Furthermore, the OCO-2 nadir and glint observations have different biases in the version 7 OCO-2 data product (the product used in this manuscript), and these differing biases make it difficult to use both types of observations in the same analysis. For example, there is a step change in the XCO_2 observations at the coastline in some locations (e.g., in parts of Africa). In these cases, the nadir mode observations may be sensitive to flux patterns, and the glint mode observations might be sensitive to flux patterns. However, an inverse model that uses both observation types together might produce unrealistic flux patterns due to the step change in XCO_2 at the coastline.

- **page 8, line 18: I would argue that the ocean is too strongly constraint by allowing only a single pattern to be adjusted in the regression. If more degrees of freedom would be assigned to the ocean, wouldn't that influence OCO-2's flux resolving performance over land?**

  We will add this caveat to the manuscript. If there are large, unresolved CO_2 fluxes from the ocean, it could influence top-down inferences of terrestrial biospheric fluxes. With that said, ocean fluxes on sub-daily time scales are much smaller than terrestrial fluxes, and the spatial patterns in these fluxes are much more diffuse than in most terrestrial regions. As a result, small errors in the distribution of marine CO_2 fluxes should not dramatically change the detectability of terrestrial fluxes.

- **page 8, line 21: this means that the biospheric flux patterns are specified per region and month, or?**

  This is correct. We tag CO_2 fluxes from each region and each month in the

PCTM atmospheric transport model. In other words, we incorporate flux patterns into PCTM at the PCTM model resolution; the model ingests CO_2 fluxes at a 1° latitude by 1.25° longitude spatial resolution and 3-hourly time resolution (Sect. 2.4 of the current ACPD manuscript). We then run the PCTM model once for each region and each month of interest. For each of these PCTM runs, we input flux patterns for the region and month of interest and zero out CO_2 fluxes for other regions and months. We will clarify that point in the associated paragraph of the revised manuscript.

- **page 8, line 27: 'stringent' in what sense? (I'd say they are rather less well contraint)**

  We agree that "stringent" is not be the best or most descriptive word here. We will replace the word "stringent" with the following phrase: "This case is more demanding of the observations than the two and four region cases; it is more difficult to obtain a robust constraint for seven regions than for two or four global regions."

- **page 8, line 31: Would this goal be achieved if the 7 biomes could be resolved by OCO- 2? Some quantitative information on how to relate surface and satellite measurements is needed here.**

  We will remove this sentence from the revised manuscript. Fang et al. (2014) examine CO_2 fluxes for North American biomes while the present ACPD manuscript focuses on global biomes. Hence, the two studies are not equivalent.

- **page 9, line 26-32: Should the reader conclude from this that we don't know whether the signal/noise analysis in figure 3 means anything?**

  We think that interpretation would be too bold. We feel that the signal/noise analysis provides useful context; it is useful to show the reader what the biospheric XCO_2 signal looks like, how it varies across the globe, and how it varies by month.

The results in Sect. 3.2 and 3.3 are based on a statistical model, and we wanted to provide an intuitive illustration of the signal and noise before presenting statistical results that use those inputs.

- **page 10, line 16: 'scales smaller than hemispheric in about half of the cases'. How can you infer information about hemispheres from a split between Tropics and Extra Tropics? The way I look at it only a single pattern is selected in 3 out of 4 seasons. Is that sufficient to resolve two pieces of information? The text suggests that OCO-2 does better than 2 ...**

  The reviewer makes a good point. A pattern is selected in approximately half of the regions and months. However, in three of the four seasons, not a single pattern is selected for one of the two hemispheres. We will add this description to the text to better represent the results.

- **page 10, line 18: 'we choose flux patterns ...' does this mean 1 or more?**

  This statement is correct. We will revise this paragraph accordingly by changing "flux patterns" to "at least one flux pattern."

- **page 10, line 32: Why is n* going down with the number of regions? Wouldn't you expect the residuals to become more random when fitting more regions? Shouldn't that make V more diagonal?**

  There are more unexplained flux patterns in the 7-region case – because model selection selects fewer variables than in the two or four region cases. As a result, the regression residuals have large covariances, and $\mathbf{V}$ is less diagonal. The variable n^* is smaller as a result.

  A brief overview of the regression helps elucidate why this is the case. The regression is iterative. We make an initial guess for n^*, run the regression with model selection, adjust n^*, and rerun the regression with model selection. We continue iterating until n^* and the regression converge – until they stop changing

from one iteration to the next. As a result, the estimate for n^* depends on which variables are included in the regression. We select a relatively small number of variables in the 7-region case, so there are many unexplained patterns in the residuals. The estimate for n^* is smaller as a result.

- **page 11, line 31: Or underestimate noise? Is there a factor in the synthetic experiments that accounts for retrieval noise?**

  The reviewer makes a great point; the estimated retrieval errors could overestimate the covariances but underestimate the variances (i.e., white noise). We will add a sentence to the paragraph explaining this point.

- **page 11, line 33: It doesn't really become clear what is mean by this "salient role". Can this be seen in the presented results?**

  This statement references the synthetic data experiments in Fig. 5. In the revised manuscript, we will specifically reference the synthetic data experiments and Fig. 5.

- **page 12, line 19: Does the relative role of transport and measurement uncertainty follow from the results of this study, or is this just speculation? It seems to me that the study should provide information on this.**

  We will clarify this result in the revised manuscript. We explore the relative roles of transport and measurement/retrieval uncertainty in the synthetic data experiments (e.g., Fig. 5 in the current ACPD manuscript). In the revised manuscript, we will make reference to the figure here and explicitly tie this statement back to the synthetic data experiments.

- **page S4, line 141: 'Consistency check'. What potential inconsistency is checked? Do you mean sensitivity or robustness check?**

  We agree that it is better to use the term "sensitivity" or "robustness" instead of "consistency." We will change the text accordingly.

- **TECHNICAL CORRECTIONS**
- **page 2, line 7: 'the the'**

  Thank you for pointing out this typo. We will correct it in the revised manuscript.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2017-813, 2017.

Characterizing biospheric carbon balance using CO_2 observations from the OCO-2 satellite

Scot M. Miller[1], Anna M. Michalak[1], Vineet Yadav[2], and Jovan M. Tadic[3]

[1]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Correspondence to: Scot M. Miller (scot.m.miller@gmail.com)

Abstract.

NASA's Orbiting Carbon Observatory–2 (OCO-2) satellite launched in summer of 2014. Its observations could allow scientists to constrain CO_2 fluxes across regions or continents that were previously difficult to monitor. This study explores an initial step toward that goal; we evaluate the extent to which current OCO-2 observations can detect patterns in biospheric CO_2 fluxes and constrain monthly CO_2 budgets. Our goal is to guide top-down, inverse modeling studies and identify areas for future improvement. We find that uncertainties and biases in the individual OCO-2 observations are comparable to the atmospheric signal from biospheric fluxes, particularly during northern hemisphere winter when biospheric fluxes are small. A series of top-down experiments indicate how these errors affect our ability to constrain monthly biospheric CO_2 budgets. We are able to constrain budgets for between two and four global regions using OCO-2 observations, depending on the month, and we can constrain CO_2 budgets at the regional level (i.e., smaller than seven global ecoregions) in only a handful of cases (16% of all regions and months). The potential of the OCO-2 observations, however, is greater than these results might imply. A set of synthetic data experiments suggests that ~~observation or~~ retrieval errors have a salient effect. Advances in retrieval algorithms and to a lesser extent atmospheric transport modeling will improve the results. In the interim, top-down studies that use current satellite observations are best-equipped to constrain the biospheric carbon balance across only continental or hemispheric regions.

1 Introduction

The OCO-2 satellite launched on July 2nd, 2014 and is NASA's first mission dedicated to observing CO_2 from space. The satellite measures the absorption of reflected sunlight within CO_2 and molecular oxygen (O_2) bands at near infrared wavelengths. These measurements are analyzed with remote sensing retrieval algorithms to yield spatially-resolved estimates of the column-averaged CO_2 dry air mole fraction, XCO_2. The satellite flies in a sun-synchronous orbit an average of 705 km above the Earth's surface, passing each location at approximately 13:30 local time, and it collects roughly 5×10^5 to 1×10^6 observations or soundings per calendar day (e.g., Crisp et al., 2004; Eldering et al., 2012; Crisp et al., 2017).

~~Unlike previous missions,~~ OCO-2 ~~observations are sensitive to CO_2 throughout the entire troposphere, advantageous~~ may provide an ideal opportunity for estimating surface CO_2 fluxes. ~~By contrast, thermal infrared observations from existing meteorological sounders such as the Atmospheric Infrared Sounder (AIRS), Tropospheric Emission Spectrometer (TES), and Infrared Atmospheric Sounding Interferometer (IASI) can yield estimates of the CO_2 concentration in the middle troposphere, but they~~ OCO-2 observes in the near-infrared, and its observations therefore have sensitivity throughout the entire troposphere with highest sensitivity near the surface (e.g., Eldering et al., 2017a). This feature contrasts with several existing satellites that observe in the thermal infrared and have little sensitivity to near-surface CO_2 variations. ~~Because~~ OCO-2 ~~observes in the near-infrared, its observations are sensitive to CO_2 variations throughout the atmospheric column, with greatest sensitivity near the surface (e.g., Eldering et al., 2017a).~~ also has a smaller footprint and improved spatial coverage relative to existing near-infrared observations like those collected by the Greenhouse Gas Observing Satellite (GOSAT). Each OCO-2 observation has a footprint ~2.25km in width, and the satellite can collect eight observations across a single swath (Eldering et al., 2017a). Each GOSAT observation, by contrast, has a footprint ~10km in width, and the satellite collects a single sounding every 250km (Yokota et al., 2009). As a result of these differences, GOSAT provides approximately 1000 high quality observations per day while OCO-2 provides approximately 65000 (e.g., Eldering et al., 2017b).

~~A few studies comment~~ Prior to the OCO-2 satellite launch, several studies commented on the possibilities of using XCO_2 observations for estimating CO_2 fluxes at the Earth's surface. For example, Chevallier et al. (2007) and Baker et al. (2010) explain that OCO-2 observations could reduce flux uncertainties by ~20–65% at the ~~the~~ model grid scale (2.5° by 3.75° and 2° by 5° latitude-longitude, respectively) and at weekly time scales. Both studies, however, caution that biases or spatially and temporally correlated errors would cut this uncertainty reduction in half. Chevallier et al. (2007) further explain that biases of a few tenths of a part per million in XCO_2 could bias estimated subcontinental flux totals by several tenths of a gigaton.

~~Relatively few studies in the published literature use XCO_2 observations from OCO-2~~ Since launch, a handful of existing studies apply the satellite observations to estimate fluxes for specific problems. Most examine flux anomalies (e.g., El Niño) or anomalously large sources (e.g., ~~primarily due to the relatively short time since the satellite's launch (e.g., Fischer et al., 2017; Heymann et a~~ ~~paucity of studies will undoubtedly change in the next several years.The literature on other CO_2 remote sensing efforts is more mature, and these studies preview the opportunities and challenge that OCO-2 may present~~power plants). For example, ~~the Greenhouse Gas Observing Satellite (GOSAT) launched in 2009 and is the first satellite dedicated to greenhouse gas monitoring~~

~~(Yokota et al., 2009).~~ Chatterjee et al. (2017), Patra et al. (2017), Heymann et al. (2017), and Liu et al. (2017) estimate flux anomalies during the most recent El Niño, and Nassar et al. (2017) estimate emissions from several large power plants.

A number of studies use GOSAT observations to estimate surface fluxes, and these studies report numerous successes and challenges that could apply to OCO-2 (e.g., Takagi et al., 2011; Basu et al., 2013; Guerlet et al., 2013; Maksyutov et al., 2013; Parazoo et al., 2013; Saeki et al., 2013; Basu et al., 2014; Deng et al., 2014; Houweling et al., 2015). GOSAT observations provide new insight into fluxes in regions that are poorly sampled by in situ observations~~, regions like tropical Asia (Basu et al., 2014) and the southern Amazon (Parazoo et al., 2013)~~. These studies also identify a number of common challenges. For example, observations are too sparse to reliably estimate fluxes in regions with frequent cloud cover (e.g., Parazoo et al., 2013). Furthermore, continental CO_2 budgets estimated using GOSAT observations are not consistent with in situ observations in some regions; these differences may indicate spatially and temporally correlated errors in GOSAT observations at their current stage of development (e.g., Houweling et al., 2015). These challenges ~~are also~~ may also be a concern for OCO-2. ~~With that said, some design features make OCO-2 even more promising than GOSAT: OCO-2 observations have a higher spatial resolution (e.g., a footprint or pixel size of about 3 km^2 relative to GOSAT's footprint of about 80 km^2) and a higher density of observations (e.g. Crisp et al., 2004; National Research Council, 2010).~~

This study evaluates the opportunities and challenges of using current OCO-2 observations to estimate biospheric CO_2 fluxes. A primary goal of this work is to guide top down, inverse modeling studies on the information content of currently-available observations. By contrast, satellite capabilities for CO_2 monitoring will likely change quickly over the next ten years – both due to improvements in satellite retrieval algorithms and the launch of new satellites (Sect. 4). This guidance will therefore undoubtedly change and evolve in the future.

We evaluate current OCO-2 observations using several approaches. We ~~make an initial assessment by comparing~~ first compare model and observation errors against the atmospheric signal from biospheric fluxes. ~~The noise in individual observations is important, but atmospheric inversions ultimately leverage broad~~ This initial comparison provides context and intuition for the signal-to-noise ratio. Ultimately, atmospheric inversions use more complex patterns in the observations ~~and other complex information~~ to estimate surface fluxes. We therefore construct a series of top-down simulations using OCO-2 observations to understand what these errors mean for estimating CO_2 fluxes, and we thereby evaluate the number of global regions for which we can independently constrain CO_2 budgets. Lastly, we construct a series of synthetic data simulations to diagnose the real data results. The first synthetic simulations do not include any errors – to evaluate the inherent strengths of the observations. In subsequent synthetic simulations, we include simulated modeling and/or retrieval errors, and evaluate how these errors affect the CO_2 flux constraint.

~~The synthetic and real data simulations are based upon a multiple regression combined with model selection. We use this approach to evaluate whether spatial and temporal patterns in biospheric CO_2 fluxes help describe patterns in the OCO-2 observations. These flux patterns are first input into a global atmospheric transport model (the Parameterized Chemistry Transport Model or are then compared against OCO-2 observations. A positive result implies that we can detect these patterns using OCO-2 observations, and a negative result implies that we cannot. Several existing studies use model selection to gauge the detectability of CO_2 flux patterns (Shiga et al., 2014; Fang et al., 2014; ASCENDS Ad Hoc Science Definition Team, 2015). The term 'patterns'~~

here refers to flux patterns that manifest at the resolution of an atmospheric model, and section 2 describes this approach in more detail.

Overall, we divide the globe into different hemispheres and ecoregions and determine whether we can detect flux patterns within each region and each month. We begin the analysis with very large hemispheric regions and then decrease the size of those regions until we are no longer able to detect any patterns beyond a mean CO_2 flux. That limit or end point is the smallest scale at which OCO-2 observations currently provide a unique constraint on CO_2 budgets. OCO-2 observations must be sufficient to detect more than a mean flux across a region and month if future inverse modeling studies are to estimate biospheric CO_2 budgets at scales smaller than that region. Consequently, inverse modeling studies would generally be unable to obtain reliable information about the fluxes across smaller regions. This result bounds the type of information one can expect from the OCO-2 retrievals in their current stage of development.

2 Methods

2.1 OCO-2 data

This study utilizes XCO_2 observations from the OCO-2 satellite beginning with the first reported data (6 Sept. 2014) through the end of 2015. We use the level 2 lite product, version B7.1.01; the lite product only includes good quality retrievals, unlike the full OCO-2 level 2 product. We only include nadir and target mode retrievals in the analysis and exclude glint mode retrievals. Recent work indicates biases in the glint retrievals relative to nadir retrievals (e.g., Wunch et al., 2017a). The SI describes model selection results with glint mode retrievals included, and the results are similar to those in the main manuscript without glint mode data.

2.2 Simulated model and retrieval errors

We simulate different model and measurement errors and compare those errors against a modeled XCO_2 signal from biospheric fluxesThe simulated model and retrieval errors provide an intuitive feel for the observations and are used as inputs in the top-down experiments later on in the study. This section describes the these simulated errors – both simulated atmospheric transport and observation or retrieval errors (Fig. 1). The SI describes these errors in greater detail.

We use estimated CO_2 transport errors from Liu et al. (2011) and Miller et al. (2015) (Fig. 1a–b). The authors of those studies run an ensemble of global meteorology simulations. CO_2 is included as a passive tracer in the model, and all simulations use the same CO_2 flux estimate but have different meteorology. The authors estimate CO_2 transport errors by examining the range of CO_2 mixing ratios in the ensemble of simulations. We choose one realization at random and subtract the mean of the ensemble from this realization to produce a set of residuals. These residuals are used as the estimated transport errors in this study. The estimated errors are therefore a realization of plausible transport errors. As a result, the specific errors used here have the same statistical properties as the other members of the ensemble but have different values in specific locations or at specific times. For example, the transport errors have a negative value across much of the Arctic in Fig. 1a. Other ensemble members , by

~~contrast,~~ might have a slight positive bias in that region but will nevertheless have similar statistical properties as the realization in Fig. 1a.

In addition to these transport errors, we simulate ~~observation or~~ retrieval errors. We use two different approaches to estimate these errors and report results using both approaches. The true retrieval errors are unknown and any effort to estimate these errors will be uncertain; the two approaches used here provide two contrasting, plausible representations of these errors.

We generate the first set of possible retrieval errors using the parameters ~~used to correct~~ in the OCO-2 ~~retrievals~~retrieval correction. This approach entails several steps. We first try to reproduce the OCO-2 observations using a regression. The predictor variables in this regression include seven different terrestrial biosphere model (TBM) estimates of net biome production and vegetation indices that have been input into an atmospheric transport model. We also include anthropogenic, ocean, and biomass burning flux estimates in the regression. A subsequent section (Sect. 2.3) describes this regression in greater detail. We save the model-data residuals from this regression. Subsequently, we regress these residuals on the parameters used in the OCO-2 bias correction. These include aerosol optical depth and albedo, among other parameters (e.g., Wunch et al., 2011, see Sect. S1.4). Note that these retrieval parameters are not run through an atmospheric transport model, unlike the ~~TBMs and vegetation indices~~TBM fluxes. We estimate the ~~observation or~~ retrieval errors as the portion of the model-observation residuals that are described by these retrieval parameters. The regression considers many different TBMs and vegetation indices, and it should therefore do reasonably well at reproducing patterns in the OCO-2 observations attributable to biospheric fluxes. Any remaining patterns in the data that map on to retrieval parameters are likely due to retrieval errors rather than transport or flux errors.

We use a second approach to create an alternative set of simulated retrieval errors. ~~We~~ Specifically, we model XCO_2 using ~~four alternative biospheric flux estimates~~ fluxes from four different TBMs (Sect. 2.3) and compute the model-data residuals for each set of simulations. We identify the grid cells in which all four sets of residuals have the same sign (i.e., generated using four different ~~flux estimates~~TBMs) and identify all of the model grid cells in which the residuals have variable sign. In the former case, we take the median of all four residuals as the estimated retrieval error and, in the latter case, assign a retrieval error of zero. This approach likely produces a conservative estimate of the retrievals errors (i.e., a possible underestimate); there is likely some amount of retrieval error at locations where we assign a retrieval error of zero.

2.3 ~~Model selection overview~~Overview of the top-down experiments

~~Model selection is a statistical approach common in regression modeling (e.g., Ramsey and Schafer, 2012, ch. 12). In many instances, a modeler must decide which predictor variablesto include (or omit) in a multiple regression . Model selection~~ We employ a top-down framework to evaluate the detectability of biospheric CO_2 fluxes using current OCO-2 observations. Overall, we divide the globe into different hemispheres and ecoregions and determine whether we can detect flux patterns within each hemisphere or ecoregion and each month. The term 'patterns' here refers to flux patterns that manifest at the resolution of an atmospheric model. We begin the analysis with very large hemispheric regions and then decrease the size of those regions until we are no longer able to detect any flux patterns beyond a mean CO_2 flux. That limit or end point is the smallest scale at which OCO-2 observations currently provide a unique constraint on CO_2 budgets. OCO-2 observations must

be sufficient to detect more than a mean flux across a region and month if future inverse modeling studies are to estimate biospheric CO_2 budgets at scales smaller than that region. Consequently, inverse modeling studies would generally be unable to obtain reliable information about the fluxes across smaller regions. This result bounds the type of information one can expect from the OCO-2 retrievals in their current stage of development. This remainder of this section explains this top-down analysis in greater detail.

We approach this problem using a regression framework. The regression attempts to reproduce OCO-2 observations of XCO_2 using predictor variables. These variables are the XCO_2 estimated by an atmospheric model; each model output estimates the XCO_2 enhancement due to fluxes in a particular region and a particular month, and we generate many model outputs using many different flux models (see Sect. 2.6). The regression has the following form:

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

where $\boldsymbol{z}$ are the OCO-2 observations (dimensions $n \times 1$). The matrix $\mathbf{X}$ ($m \times p$) has p columns, and each column is a different CO_2 flux estimate for specific geographic region and a specific month. Each column of $\mathbf{X}$ has non-zero values for a specific region and month and zeros for all other regions and months. The function $h()$ is an atmospheric transport model (Sect. 2.6), and the model outputs, $h(\mathbf{X})$, have dimensions $n \times p$. The vector $\boldsymbol{b}$ ($n \times 1$) is the model spin-up or estimated XCO_2 at the beginning of the study time period, and $\boldsymbol{\epsilon}$ ($n \times 1$) are the regression residuals. Lastly, $\boldsymbol{\beta}$ ($p \times 1$) are the coefficients estimated as part of the regression.

The regression provides a means to evaluate the detectability of biospheric CO_2 fluxes. At least some of the model outputs ($h(\mathbf{X})$) should describe substantial variability in the OCO-2 observations ($\boldsymbol{z}$). Let's say that modeled XCO_2 using a particular flux model in a particular region and month help reproduce patterns in the OCO-2 observations. This result implies that OCO-2 observations are able to detect or constrain variability in CO_2 fluxes from that region and that month. By contrast, let's say that no model outputs substantially improve the regression fit (i.e., no columns in $h(\mathbf{X})$). This result implies one of three things. First, the OCO-2 observations may not be sensitive to biospheric fluxes from that particular region and month. Second, there may be errors in current OCO-2 retrievals or in the atmospheric model that obscure surface fluxes from that region and month. Finally, all of the flux estimates used in $\mathbf{X}$ may be unskilled and may not match real-world conditions. We offer up a large number of flux models as possible predictor variables in $\mathbf{X}$, and at least some of these products should be expected to correlate with real world CO_2 fluxes. In this case, it is unlikely that there is a shortage of reasonable CO_2 flux models to choose from. Rather, that result more likely reflects the sensitivity of the observations to surface fluxes, the maturity of current OCO-2 retrievals, or the accuracy of the atmospheric model. Hence, this approach provides a means to evaluate when and where current OCO-2 observations can constrain variability in biospheric CO_2 fluxes.

We utilize a model selection framework to determine which model outputs (i.e., columns of $h(\mathbf{X})$) describe substantial variability in current OCO-2 observations. Model selection is a statistical approach common in regression modeling (e.g., Ramsey and Scha
It will identify the set of predictor variables with the greatest power to describe the data. It also ensures that the regression does not overfit the data. The inclusion of more predictor variables in a regression will always improve model-data fit; a regression with n independent predictor variables will always be able to describe n data points perfectly. However, a model with n inde-

pendent predictor variables would overfit the data~~(for more on the dangers of overfitting, refer to Zucchini, 2000).~~ . For more on the dangers of overfitting, refer to Zucchini (2000). To this end, one can use model selection to prevent overfitting and only include predictor variables that describe substantial variability in the ~~data~~observations.

We implement model selection based on the Bayesian Information Criterion (BIC) (Schwarz, 1978). We calculate a BIC score for many different combinations of predictor variables, and each combination has a different set of columns ($\mathbf{X}$). The best combination has the lowest BIC score:

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

where L is the log likelihood of a particular combination of predictor variables (i.e., a particular configuration of $\mathbf{X}$). The log likelihood equation rewards combinations that improve model-data fit, and the second term in the equation ($p\ln n^*$) penalizes combinations with a greater number of predictor variables (i.e., columns in $\mathbf{X}$). This penalty also scales with the effective number of independent observations from OCO-2 (n^*, described in the SI), and it ensures that the selected model is not an over-fit.

A number of existing top-down studies of CO_2 use model selection (e.g., Gourdji et al., 2008, 2012; Shiga et al., 2014; Fang et al., 2014; Fang and Michalak, 2015; ASCENDS Ad Hoc Science Definition Team, 2015). Several ~~use~~ utilize the approach to determine a set of environmental variables to include in a geostatistical inverse model (e.g., Gourdji et al., 2008, 2012). Other studies use model selection to determine whether existing CO_2 observations can constrain flux patterns from the biosphere (Fang et al., 2014) and from fossil fuel emissions (Shiga et al., 2014; ASCENDS Ad Hoc Science Definition Team, 2015). One study uses model selection to assess the capabilities of a proposed satellite mission (ASCENDS Ad Hoc Science Definition Team, 2015).

~~We use model selection to explore whether current OCO-2 observations are sufficient to detect broad spatial and temporal patterns in CO_2 fluxes – within two large hemispheric regions, four continental regions, and seven ecoregions (Fig. 2). The last goal is more challenging than the first. In each case, we use model selection to examine whether the flux patterns within each region and each month help reproduce patterns in the OCO-2 observations.~~

~~We begin with a baseline model that has spatial and temporally constant fluxes for each region and month of interest. These constant terms are analogous to the intercept in a multiple regression. These intercept terms are always included (e.g., Gourdji et al., 2008; Fang et al., 2014), and we use model selection to identify additional model outputs as necessary.~~

~~We subsequently model XCO_2 using an atmospheric transport model (Sect. **??**) and several different biospheric flux estimates and vegetation indices. We then incorporate these model outputs as predictor variables in a regression and use model selection to identify which patterns (if any) explain substantial variability in the OCO-2 observations. These patterns include four TBMs with contrasting spatial features from MsTMIP, the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (Huntzinger et al., 2013; Fisher et al., 2016; **?**). Section S1.2 describes MsTMIP and the TBMs in greater detail. We also include SIF (solar-induced fluorescence) from the Global Ozone Monitoring Experiment-2 (GOME-2, Joiner et al., 2013) as well as EVI (enhanced vegetation index) and NDVI (normalized difference vegetation index) from the Moderate-Resolution~~

Imaging Spectroradiometer (MODIS; e.g., Huete et al., 2002). Note that we directly input these vegetation indices into an atmospheric transport model as a surface 'flux.' The regression/model selection framework will adjust the magnitude of the transport model outputs to reproduce the OCO-2 observations, so the absolute magnitude of the vegetation indices is not important. Rather, we are interested in whether the patterns in these vegetation indices help reproduce patterns in the OCO-2 observations, potentially in combination with other indices or TBMs.

We offer up a relatively large number of flux models and vegetation indices as predictor variables, and at least some of these products are therefore expected to correlate with real world CO_2 fluxes. We should choose at least one of these variables using model selection if the OCO-2 observations are able to detect patterns in the surface fluxes. If we do not choose any additional outputs with model selection, it suggests that the observations are not sufficient to detect spatial and temporal patterns in the fluxes beyond a mean flux. We include a large number of candidate variables for a pragmatic reason. If model selection does not pick any variable in a region, it is unlikely that there was a shortage of reasonable CO_2 flux patterns available to choose from. Rather, that result more likely reflects the maturity of current OCO-2 observations and atmospheric modeling capabilities.

Model selection Model selection provides a convenient way to evaluate the information content of OCO-2 observations in their current state of development. In theory, one could estimate CO_2 budgets in a Bayesian inverse model. The accuracy or uncertainty in those budgets would be indicative of the information content of the satellite observations. This approach, however, brings several challenges. First, a modeler must choose a prior flux estimate. This choice is often subjective but will impact the final or posterior uncertainty estimate (e.g., Chevallier et al., 2014). Second, a modeler must estimate several individual sources of uncertainty as inputs to the inverse model. These uncertainties often have a complex statistical structure that is difficult to characterize (e.g., Liu et al., 2011), and it is often challenging to account for all plausible sources of uncertainty. Third, inverse modeling with satellite observations can be computationally intensive – both in terms of the number of atmospheric model simulations required and the computational requirements of the statistical inverse model. Some studies have overcome the first of these two challenges by using an ensemble of atmospheric models and/or inversion systems (e.g., Chevallier et al., 2014; Houweling et al., 2015). The size of the ensemble spread is indicative of the information content of the observations, and the spread of the ensemble is usually larger than the uncertainty bounds estimated from any one inverse model. This type of study also typically requires extensive coordination among multiple research groups. Model selection, by contrast, By contrast, the regression framework used here provides a simpler metric to evaluate the information content of the observations.

The remainder of this sub-section describes the specific equations used for model selection . We quantitatively link the OCO-2 XCO_2 observations with model outputs using a multiple regression :

$$\mathbf{Y} = h(\mathbf{X})$$
$$\boldsymbol{z} = \mathbf{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V})$$

In these equations, the vector $\boldsymbol{z}$ (dimensions $n \times 1$)represents the XCO_2 observations minus the model initial condition or spin-up (refer to the SI for more detail). The variable

2.4 Implementation of the top-down experiments

This section describes how the regression and model selection are implemented in the present study.

The regression begins with an intercept. The intercept is always included in the regression (in $\mathbf{X}$), and model selection can further add flux models to $\mathbf{X}$ ~~(dimensions $m \times p$)is a matrix of p different flux patterns; the columns of $\mathbf{X}$ can include biospheric CO_2 flux estimates, remote sensing vegetation indices, an anthropogenic emissions inventory, a biomass burning inventory, and /or an ocean fluxestimate. The function $h()$ is an atmospheric model that transports the fluxes to the times and locations of the~~ to help reproduce the OCO-2 ~~retrievals, and the resulting matrix $\mathbf{Y}$ has dimensions $n \times p$. The variable $\boldsymbol{\epsilon}$ is a $n \times 1$ vector of residuals. These residuals are assumed to follow a multivariate normal distribution with a mean of zero, a variance of σ^2, and a covariance structure given by $\mathbf{V}$ (dimensions $n \times n$). The vector of coefficients ($\boldsymbol{\beta}$, dimensions $p \times 1$) are estimated as part of the regression .~~

~~In this study, we choose a set of variables for~~ observations ($\boldsymbol{z}$). This intercept is a constant column of ones in $\mathbf{X}$~~using model selection based on the Bayesian Information Criterion (BIC) (Schwarz, 1978). We calculate a BIC score for many different candidate models. Each candidate model has a different set of columns ($\mathbf{X}$)~~ . In the particular setup here, we include a different intercept for each region of the globe and each month. In other words, the intercept consists of multiple columns – ~~different combinations of flux models or remote sensing vegetation indices in different geographic regions~~one column for each region and month of the study period. This intercept is equivalent to a spatially and temporally constant flux in each region and month. Additional atmospheric model outputs $h(\mathbf{X})$ will not be selected unless they explain substantially more variability in the OCO-2 observations than this intercept or constant flux. The intercept plays an important role in the regression; it ensures that the regression will always be unbiased when averaged across the globe. Model selection could produce non-intuitive results if there were no intercept. Furthermore, intercepts are standard in regression modeling, and ~~in different months.~~

~~The best model has the lowest BIC score:~~

$$BIC = L + p\ln(n^*)$$

~~where L is the log likelihood of a particular candidate model ($\mathbf{X}$~~top-down, CO_2 studies that utilize model selection also include an intercept (e.g., Gourdji et al., 2008, 2012; Fang et al., 2014)

We subsequently run the regression with model selection three times to evaluate three different cases. In the first case, we divide the flux models into two large hemispheric regions, and we select different flux models for each hemispheric region and each month. The second and third experiments divide the fluxes into four continental regions and seven ecoregions, respectively (Fig. 2). The ~~log likelihood equation rewards models that improve model-data fit, and the second term in the equation ($p\ln n^*$) penalizes complex models; it ensures that the selected model is not an over-fit. The log-likelihood has the following form:~~

$$\begin{aligned} L &= n^* \ln(\sigma^2) + \tfrac{n^*}{n} RSS \\ RSS &= \tfrac{1}{\sigma^2}\boldsymbol{z}^T\boldsymbol{z} - \tfrac{1}{\sigma^2}\boldsymbol{z}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\boldsymbol{z} \end{aligned}$$

~~where RSS is the residual sum of squares and σ^2 is defined above in Eq. 3.~~ last experiment is more challenging than the first.

~~Both the BIC and log-likelihood equations (Eq. 2 and 3) incorporate n^*, the effective number of independent observations~~

Note that we consider the model selection results from 2014 part of an initial model spin-up period and only report the results from 2015.

2.5 Synthetic data simulations

We subsequently utilize synthetic data simulations in this study to analyze the effects of different model or retrieval errors on the detectability of biospheric CO_2 fluxes. In these simulations, we create synthetic XCO_2 observations using an atmospheric transport model (Sect. 2.6) and the SiBCASA flux model. We then run model selection using these synthetic observations in place of real OCO-2 observations ($\boldsymbol{z}$). These model selection runs have the same setup as the real data simulations, except that the observations are synthetic instead of real. Furthermore, we only analyze the seven ecoregion case in the synthetic data experiments. This case is more demanding of the observations than the two and four region cases; it is more difficult to obtain a robust constraint for seven regions than for two or four larger global regions. The seven ecoregion case is also an important goal from a ecological perspective. For example, one might want to estimate how CO_2 fluxes differ in different tropical forests or in different temperate forests (e.~~Jones (2011) discusses this concept in the context of the BIC. Just because the satellite provides n observations does not mean there are n independent pieces of information. Accordingly, n^* ensures that the model selection framework accurately assesses the amount of independent information in the observations . It accounts for the fact that there are often spatial and temporally coherent errors in the satellite observations or in the transport model. If all of the observations were independent (i.e., if $\mathbf{V}$ were diagonal), then n^* would equal n.However, we de-weight both components of Eq. 2 as the covariances in $\mathbf{V}$ increase~~g., on different continents or in different climate zones).

~~We could calculate n^* directly using $\mathbf{V}^{-1}$ (Jones, 2011). In fact, several~~ These synthetic simulations help to isolate the effect of different errors on the detectability of biospheric CO_2 ~~model selection studies incorporate $\mathbf{V}^{-1}$ directly into the equation for RSS (e.g., Mueller et al., 2010; Gourdji et al., 2012; Shiga et al., 2014). We use 5,079,165 observations (n)in this study, so $\mathbf{V}$ has 5.08×10^6 rows and columns. As a result, the inverse of $\mathbf{V}$ is computationally intractable. We instead estimate n^* using an approach based on Griffith (2005), an approach that does not require computing $\mathbf{V}^{-1}$ directly:~~

~~$$n^* = \frac{n}{1 + \left(\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} V_{i,j}/n\right)}$$~~

~~We estimate $V_{i,j}$ in the vicinity of each observation i by fitting a local variogram model on the model-data residuals ($\boldsymbol{\epsilon}$) (See Alkhaled et al. (2008) and Hammerling et al. (2012) for a description of local variogram analysis.). The SI describes this implementation in greater detail.~~

2.6 ~~OCO-2 data~~

~~This study utilizes XCO~~fluxes. We first run the regression and model selection with no errors in the atmospheric model or in the retrievals ($\boldsymbol{\epsilon} \approx \mathbf{0}$). We then successively add simulated error to the synthetic observations and evaluate how the model selection results change as the errors increase (Fig. 1). We include three different types of errors: flux errors, transport errors,

and retrieval errors. The simulations with all errors included should produce model selection results similar to the real data experiments. Section 2.2 and the SI describe the simulated transport and retrieval errors.

The flux errors further account for plausible inaccuracies in the predictor variables within $\mathbf{X}$. No TBM or vegetation index has a distribution that perfectly matches real-world CO_2 ~~observations from~~ fluxes. These errors affect our ability to detect biospheric fluxes using the OCO-2 ~~satellite beginning with the first reported data (6 Sept. 2014) through the endof 2015. We use the level 2 lite product, version B7.1.01; the lite product only includes good quality retrievals, unlike the full OCO-2 level 2 product. We only include nadir and target mode retrievals in the analysis and exclude glint mode retrievals. Recent work indicates biases in the glint retrievals relative to nadir retrievals (e.g., Wunch et al., 2017b). The SI describes model selection results with glint mode retrievals included, and the results are similar to those in the main manuscript without glint mode data~~observations, and we therefore account for these flux errors in the synthetic data simulations. To this end, we remove SiBCASA as an option in the $\mathbf{X}$ matrix and choose among the other remaining flux models. This procedure simulates the plausible effects of imperfect flux models or predictor variables.

2.6 Atmospheric ~~transport~~ model simulations

We employ ~~PCTM~~ the Parameterized Chemistry Transport Model (PCTM) to model XCO_2 using a variety of surface flux ~~estimates and vegetation indices~~ models (Kawa et al., 2004). A number of existing studies use this model to simulate atmospheric CO_2 mixing ratios (e.g., Law et al., 2008; Gurney et al., 2009; Baker et al., 2010; Schuh et al., 2010; Shiga et al., 2013; ASCENDS Ad Hoc Science Definition Team, 2015; Hammerling et al., 2015). Several of these studies specifically use PCTM to model CO_2 in the context of satellite missions (e.g., Baker et al., 2010; ASCENDS Ad Hoc Science Definition Team, 2015; Hammerling et al., 2015). The PCTM configuration in this study has global coverage, a spatial resolution of 1° latitude by 1.25° longitude, and 56 vertical levels. We both input the fluxes and estimate atmospheric mixing ratios at a 3-hourly time resolution, and the model transports fluxes through the atmosphere using winds from NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Rienecker et al., 2011). Section S1.1 includes more detail on the model initial condition and spin-up period.

2.7 ~~Real and synthetic data simulations~~

~~We initially use real observations to examine whether current OCO-2 observations can detect biospheric flux patterns in each month and each region of the globe.~~ We ~~consider the model selection results from 2014 part of an initial model spin-up period and only report the results from 2015.~~ subsequently model XCO_2 using several different terrestrial biosphere models (TBMs) and vegetation indices, and these model outputs are incorporated into model selection ($h(\mathbf{X})$). We include four TBMs with contrasting spatial features from MsTMIP, the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (Huntzinger et al., 2013; Fisher et al., 2016; **?**). Section S1.2 describes MsTMIP and the TBMs in greater detail. We also include SIF (solar-induced fluorescence) from the Global Ozone Monitoring Experiment-2 (GOME-2, Joiner et al., 2013) as well as EVI (enhanced vegetation index) and NDVI (normalized difference vegetation index) from the Moderate-Resolution Imaging Spectroradiometer (MODIS; e.g., Huete et al., 2002). Note that we directly input these vegetation indices into PCTM

as a surface 'flux.' The regression will adjust the magnitude of the transport model outputs to reproduce the OCO-2 observations, so the absolute magnitude of the vegetation indices is not important. Rather, we are interested in whether the patterns in these vegetation indices help reproduce patterns in the OCO-2 observations, potentially in combination with other indices or TBMs.

We also consider non-biospheric fluxes for use in $\mathbf{X}$. We include anthropogenic emissions from EDGAR v4.2 FT2010 (European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL), 2013; Olivier et al., 2014), climatological ocean fluxes from Takahashi et al. (2016), and biomass burning fluxes from the Global Fire Emissions Database (GFED), version 4.1 (van der Werf et al., 2010; Giglio et al., 2013). We are not interested in anthropogenic or marine fluxes per se. Rather, we want to account for these fluxes in the modeling framework and do not want any omissions to affect inferences related to biospheric fluxes. As a result, we do not separate these non-biospheric fluxes by region or month because these sources are not the focus of this study; each of these sources is assigned a single column in $\mathbf{X}$. Furthermore, we do not include these variables by default within $\mathbf{X}$, unlike the constant flux base model. Rather, they are included as candidate variables within the model selection framework.

~~In the seven~~ Note that in the seven ecoregion region case, $\mathbf{X}$ has a minimum of 112 columns and a maximum of 899 columns~~—~~. 112 columns associated with the ~~constant flux base model,~~ intercept and are always included in $\mathbf{X}$ (i.e., 16 months $\times$ 7 ecoregions per month). 784 columns associated with biospheric ~~flux patterns~~fluxes (7 flux models $\times$ 16 months $\times$ 7 ecoregions), and three columns associated with fossil fuel, ocean, and biomass burning fluxes. These columns may or may not be included in $\mathbf{X}$, depending upon the results of model selection.

~~We further utilize synthetic data simulations in this study to evaluate the potential of OCO-2 observations and analyze the effects of different model or retrieval/observation errors. In these simulations, we create synthetic XCO_2 observations using PCTM and the SiBCASA model. We then run model selection using these synthetic data. These model selection runs have the same setup as the real data simulations, except that the observations ($\boldsymbol{z}$) are synthetic instead of real. Furthermore, we only analyze the seven region case in the synthetic data experiments. This case is more stringent than the two and four region cases. The seven ecoregion case is also an important goal from a ecological perspective: one might want to estimate how CO_2 fluxes differ in different tropical forests or in different temperate forests (e.g., on different continents or in different climate zones). The in-situ monitoring network in North America, for example, is able to detect flux patterns within most North American ecoregions (Fang et al., 2014), and this standard of detectability would be a rigorous goal for satellites to achieve.~~

~~The first model selection runs are an idealized case with no simulated errors ($\boldsymbol{\epsilon} \approx \mathbf{0}$). We then successively add simulated error to the synthetic observations and evaluate how the model selection results change as the errors increase (Fig. 1). We include three different types of errors: flux errors, transport errors, and observation or retrieval errors. Section 2.2 and the SI describe the simulated transport and retrieval errors. The flux errors further account for plausible inaccuracies in the predictor variables within $\mathbf{X}$. No TBM or vegetation index has a distribution that perfectly matches real-world CO_2 fluxes. These errors affect our ability to identify biospheric flux patterns using the OCO-2 observations and would affect inverse modeling studies; the better the prior flux estimate in an inverse model , the more accurate and reliable the posterior flux estimate. We therefore account for these flux errors in the synthetic data simulations. To this end, we remove SiBCASA as an option in the $\mathbf{X}$ matrix~~

and choose among the other remaining patterns. This procedure simulates the plausible effects of imperfect TBMs or predictor variables.

3 Results & discussion

3.1 The biospheric CO_2 signal versus model and measurement noise retrieval errors

We compare simulated model and measurement retrieval errors against the atmospheric signal from biospheric fluxes (Fig. 3). The comparison provides an intuition of the 'noise' errors and the CO_2 'signal' given current modeling and observation capabilities. This Prior to the OCO-2 satellite launch, several studies modeled the XCO_2 signal from surface fluxes (e.g., Olsen and Randerson, 2004), and the measurement precision required for space-based constraints on CO_2 fluxes (e.g., Rayner and O'I
It is now possible to make this evaluation with real instead of synthetic observations. Also note that the XCO_2 signal and estimated errors will vary depending on the averaging time period. We report monthly averages of the biospheric signal and the errors. Many top-down, inverse modeling studies report monthly flux totals, so all of the analysis presented here is aggregated to one month averages.

The atmospheric CO_2 signal from biospheric fluxes is marked, even when averaged across a total vertical column . Globally, the XCO_2 signal has a (Fig. 3)– a global mean absolute value of 0.5ppm in February and 1.3ppm in July. The 10th and 90th percentiles are 0.04 and 1.4ppm in February and 0.06 and 3.8ppm in July. In July, the largest enhancements are in the northern hemisphere mid and high latitudes while the largest enhancements during winter months are in the tropics and southern hemisphere.

The By contrast, the simulated model and retrieval /observation errors are comparable to this XCO_2 signal from biospheric fluxes. These errors have a mean absolute value of 0.6ppm in both February and July. The 10th and 90th percentiles are 0.08 and 1.35ppm in February and 0.08 to 1.25ppm in July. Using the alternative retrieval estimate, the errors and percentiles are 0.8, 0.02, and 2.1ppm in February and 1.4, 0.05, and 3.7ppm in July. Phrased differently, these errors, averaged across all nadir and target data, equate to 115 – 122% of the mean biospheric CO_2 signal in February and 43 – 107% in July, depending upon the retrieval error estimate. Figure 3 further indicates the spatial distribution of the atmospheric CO_2 'signal' and 'noise.'

It is important to note that the distribution of these observations is heterogeneous across the globe (Fig. 3), even though the total number of OCO-2 observations is large (e.g., 268,671 and 343,053 nadir and target observations in February and July, respectively, in the lite data file), but the distribution of these observations is heterogeneous across the globe. For example, the data are concentrated in tropical and temperate regions and are sparse at high latitudes and regions with frequent cloud cover (e.g., the Amazon).

These errors are not inconsistent with those estimated by existing studies. Wunch et al. (2017a) and Liang et al. (2017) compare OCO-2 observations against TCCON observations. Wunch et al. (2017a) report an average site bias of 0.22ppm when using land nadir retrievals and a root mean squared error of 1.31ppm. By contrast, Liang et al. (2017) report an error standard deviation of 1.56ppm (both nadir and glint), and find that OCO-2 observations are offset from GOSAT observations by an average of 1.765ppm. In addition, Worden et al. (2017) estimate a precision and accuracy, respectively, of 0.75 and 0.65ppm

for land nadir observations. These studies are not necessarily directly comparable because each uses different metrics and observations. Furthermore, several compare OCO-2 to TCCON, the same observations used to bias correct OCO-2. With that said, the overall numbers reported by different studies are not inconsistent with one another and with Fig. 3.

The relative magnitude of the errors provides an informative measure of the observations, but it does not tell the complete story. A number of other considerations affect scientists' ability to estimate surface fluxes using these observations. First, inverse models leverage more than the point-wise signal to estimate surface fluxes; these models leverage complex spatial and temporal patterns in the data to estimate surface fluxes. Second, the absolute magnitude or variance of the errors is only one consideration. Another important factor is the spatial and temporal correlations or covariances in these errors. These covariances reduce the independent information in the data and can obscure patterns in XCO_2 that are due to surface fluxes. As a result, we construct real and synthetic data experiments to understand what these errors mean for estimating surface fluxes.

3.2 Real data experiments

The model selection experiments using real data indicate the number or size of regions for which we can reliably constrain biospheric CO_2 budgets using current OCO-2 observations. We start the real data simulations with large hemispheric regions and reduce the size of the regions until we are no longer able to detect any CO_2 flux patterns or information beyond a mean monthly flux in each region and each month. We would need to detect more than a mean flux from a given region if we are to reliably constrain fluxes across smaller regions.

The first real data experiment indicates whether the observations are sufficient to detect flux patterns within two large hemispheric regions (Fig. 2). Figure 4a displays the number of months in which at least one ~~pattern~~ model output (i.e., column of $\mathbf{X}$, Eq. 1) is chosen, broken down by region. The results in Fig. 4a are grouped by season for convenience. ~~Dark colors suggest excellent detectability for a given region and season while light colors suggest limited detection abilities.~~

Model selection identifies flux patterns in about half of all months. This outcome suggests that OCO-2 and the PCTM model can be used to identify broad, hemispheric flux patterns. One important exception is the extra-tropics (e.g., the temperate, boreal, and arctic region), in both spring and fall. Biospheric uptake in these seasons is less than the summer maximum, in both the northern and southern hemispheres. As a result, flux patterns in these areas are not as heterogeneous and not readily detectable using the satellite observations. This result also indicates that the OCO-2 observations can be used to reliably constrain CO_2 budgets at scales smaller than hemispheric in about half of all cases. With that said, we do not select a single flux pattern in three of four seasons for one hemisphere.

In a second experiment, we try to identify flux patterns within four, smaller regions using OCO-2 observations and model selection (Fig. 4b). ~~We choose flux patterns~~ At least one flux pattern is selected in 29% of all regions and months~~using model selection~~, and this result suggests that inverse modeling studies would be able to constrain CO_2 budgets at more detailed spatial scales in about one third of all regions and months. This experiment is more demanding than the first, and it is therefore unsurprising that ~~we choose~~ fewer flux patterns ~~; flux~~ are selected. Flux patterns within these four continental regions are often less heterogeneous than across the two large hemispheric regions in the first experiment. ~~This result suggests that inverse~~

~~modeling studies would be able to constrain CO_2 budgets at more detailed spatial scales in about one third of all regions and months.~~

The third and final experiment includes seven ecoregions, and ~~we choose~~ relatively few flux patterns ~~using model selection~~ are selected in this final experiment (16% of all possible regions and months, Fig. 4c). This result suggests a limited ability to detect biospheric flux patterns within each of the seven global ecoregions. Inverse modeling studies would thus be able to uniquely constrain CO_2 budgets across smaller regions in only a small handful of cases (e.g., 16% of all possible regions and months). These results appear similar to a recent study using current GOSAT retrievals. Houweling et al. (2015) compare an ensemble of inverse modeling flux estimate using GOSAT. Estimates show good agreement across very large regions (e.g., within 20% for global, annual CO_2 budgets) but disagree by over 100% over subcontinental-sized TransCom regions (e.g., Gurney et al., 2002).

As part of model selection, we also ~~estimate~~ evaluate the effective number of independent OCO-2 observations ~~, referred to as n^* (Eq. S7)~~(n^*, Sect. S1.3), and we estimate one independent observation per ~1200 OCO-2 lite retrievals. Correlations or covariances in transport and retrieval errors will reduce the value of n^*. ~~Our~~ The estimate is similar among all of the real data experiments and is approximately 4000 (4060 for the two region case, 3600 for the four region case, and 3540 for the seven region case). By comparison, the total number of OCO-2 observations during the study period (n) is 5.08×10^6. ~~This ratio corresponds to one independentobservation per ~1200~~ Note that the value of n^* decreases as the number of regions increases. Fewer and fewer model outputs are selected as the number of regions increases. As a result, there are more residual, unexplained flux patterns in the seven region case than in the two region case. As a result, the regression residuals have larger covariances in the seven region case, and the observations become less independent.

These results are broadly consistent with the OCO-2 ~~lite retrievals.~~ science team's ongoing flux inter-comparison study (**?**). Several research groups are developing inverse models to estimate CO_2 fluxes using OCO-2 observations, and a comparison of these estimates provides insight into the robustness of the flux estimates. To date, these comparisons often show relatively good agreement for total hemispheric terrestrial budgets, but the ensemble of estimates of diverges for smaller regions. Other, newly published inverse modeling studies use OCO-2 observations to estimate regional budgets (e.g., Liu et al., 2017). These studies primarily focus on questions about carbon cycle science, and many employ 4DVAR frameworks that are not necessarily intended to exhaustively sample all plausible sources of uncertainty.

3.3 Synthetic data experiments

The goal of the synthetic data simulations is to understand the challenges that influence the real data results. If future efforts can mitigate these challenges, then inverse modeling studies would be able to reliably constrain flux patterns and CO_2 budgets across individual ecoregions or even smaller regions.

We first construct an idealized synthetic data study without any errors (Fig. 5a)~~, and we choose a flux pattern for every ecoregion in every month using model selection~~. This case study indicates that the OCO-2 observations are not inherently insensitive to biospheric fluxes at the surface~~. This result is not necessarily an obvious one. Some satellite instruments, particularly those that measure in the thermal infrared, are most sensitive to mixing ratios in the upper troposphere and have limited~~

~~sensitivity to the surface (e.g., AIRS, see Chevallier et al., 2005). OCO-2, by contrast, observes in the shortwave infrared and is sensitive to CO_2 mixing ratios throughout the troposphere.~~ , a result consistent with previous studies (e.g., Olsen and Randerson, 2004). In this idealized case, a flux pattern is selected for every ecoregion in every month using model selection.

Subsequent model selection experiments include at least one error type, and ~~these results all look different from the first, idealized case. We choose patterns~~ flux patterns are selected in fewer regions and months ~~with model selection when we include one error type~~ in all of these cases (Fig. 5b-d). Of the three different error types, ~~observation or~~ retrieval errors have the largest impact on the model selection results (Fig. 5d). Note that the retrieval errors used to generate Fig. 5d are those simulated in Fig. 1c-d. Section S2.2 describes the model selection experiments using an alternative set of retrieval errors (Fig. 1e-f), and these results are similar to those presented in the main manuscript (Fig. 5).

Notably, the addition of any error, large or small, appears to hinder flux pattern detection in marginal ecoregions – ecoregions with small fluxes and/or small spatial and temporal variability (e.g., tundra and deserts). Arguably, this result is unsurprising. OCO-2 observations are sparse in many cloudy, high latitude regions, and CO_2 fluxes are weak at high latitudes and in deserts. Fluxes from these regions are quickly obscured by even modest errors in the model or observations. Future CO_2 remote sensing efforts would have difficulty detecting biospheric patterns within these areas. Other regions, like forests and grasslands, have larger and/or more heterogeneous fluxes, and these patterns should be easier to detect with satellite observations.

~~The experiments in~~ One of the experiments specifically accounts for errors in existing flux estimates (Fig. 5b). We find that these errors do have an effect on the results, but that effect is not nearly as large as that due to retrieval errors. We argue that current OCO-2 observations can detect patterns in CO_2 fluxes if at least one model output helps explain substantial variability in those observations. We offer up a number of different model outputs using several different flux models, but there is always a possibility that none of these flux models adequately approximates real world fluxes. The experiment in Fig. 5b evaluates how these flux errors could affect the result. Furthermore, if there are large, unresolved patterns in anthropogenic or marine fluxes, these unresolved patterns could also influence the results for biospheric fluxes. Note that this set of issues also affects Bayesian inverse models and therefore has implications beyond the methodology used in this study. Specifically, the availability and skill of the prior flux estimate affects the robustness and uncertainty of the inverse modeling estimate.

Subsequent experiments (Fig. 5e-g) include two different error types, and ~~we choose~~ flux patterns are selected in fewer months and regions ~~with model selection~~ relative to the previous cases. The experiment with transport and flux errors (but not retrieval errors, Fig. 5e) still show good detectability during the summer months and in temperate and tropical forest and grassland ecoregions. By contrast, the experiments that include retrieval errors (Fig. 5f-g) show limited detectability in all ecoregions and seasons.

The last experiment (Fig. 5h) includes all error types~~. We~~ , and we obtain positive results in fewer regions in fewer months relative to other cases. These results are broadly consistent with the real data experiments; ~~we choose~~ a similar number of regions and seasons are selected in both this experiment and the real data experiment (Fig. 5h). This consistency indicates that the synthetic simulations likely mirror real-world conditions. Note that the estimate for n^* in this final synthetic experiment is about half that of n^* in the real data experiments ($n^* = 1630$). These synthetic experiments may therefore slightly overestimate

the ~~spatial and temporal covariance~~ error correlations or covariances and underestimate the variance or white noise portion of the errors~~relative to real data~~.

Overall, the synthetic simulations suggest that ~~observation or~~ retrieval errors play a salient role relative to other error types (e.g., transport errors or flux errors, Fig. 5). Spatial and temporal error covariances and biases may be at least partly to blame. The estimated transport errors are spatially and temporally correlated on synoptic time scales ~~(e.g., Miller et al., 2015)~~(e.g., Liu et al., 2011; These scales are generally smaller than the ecoregions and hemispheres examined in this study. As a result, these errors will average down over time and space, and this averaging will mitigate the impact of these errors on the results. This statement, of course, only holds if there are no large-scale biases in the meteorology. The simulated ~~observation or~~ retrieval errors, by comparison, covary across longer spatial and temporal scales. These errors correlate with retrieval parameters like aerosol optical depth or albedo that often change at broader seasonal or regional scales. The greater these error correlations, the less these errors average down across space and time, and the greater impact these errors have on the utility of XCO_2 observations. A reduction in the spatial and temporal coherence of these errors would improve the model selection results.

4 Conclusions

The OCO-2 satellite offers a new, global window into atmospheric CO_2 and CO_2 fluxes at the Earth's surface. This study explores a first step in realizing these capabilities; we evaluate the extent to which current OCO-2 observations can detect patterns in biospheric CO_2 fluxes and constrain monthly CO_2 budgets.

We find that OCO-2 observations, in their current state of development, often provide a reliable constraint on CO_2 budgets across continental or hemispheric regions. By contrast, we find that current observations can provide a unique CO_2 estimate across smaller regions in only a handful of cases. As a result, inverse modeling studies are unlikely to constrain regional fluxes at fine spatial and temporal scales given the current maturity of the observations. Regional CO_2 budgets estimated using these observations would be highly uncertain and prone to biases.

These results do not reflect any inherent limitation in the sensitivity of the OCO-2 satellite. Rather, ~~they~~ a set of synthetic data simulations indicate that these limitations are likely the ~~product of observation or~~ result of large errors: retrieval errors and to a lesser extent atmospheric transport errors ~~.~~ (Fig. 5). Hence, the value or potential of the OCO-2 observations is greater than these results might otherwise imply. For example, ~~our simulated retrieval errors~~ the retrieval errors simulated in this study often covary across large regions and across a month or more. Future improvements to retrieval algorithms could reduce both the variance and covariance of these errors, enabling confident CO_2 flux constraints across smaller regions.

Even with these limitations, current OCO-2 observations provide new information on CO_2 fluxes for many regions of the globe. On one hand, in situ data appear to provide a stronger constraint on CO_2 fluxes in some well-instrumented regions of the world, like North America (e.g., Fang et al., 2014). ~~Our results using OCO-2 and~~ Results using seven global ecoregions ~~are less successful~~show only a limited ability of current OCO-2 observations to differentiate among regions. On the other hand, in situ observations are sparse in many regions of the world, including in most of the tropics, Africa, South America, and Asia

(e.g., **?**). Current OCO-2 observations bring new monitoring capabilities to these regions that are unlikely to be matched by in situ observations within the near future.

Furthermore, a number of new satellite missions will launch in the next five years. Multiple sets of observations, in tandem, will provide a more detailed, robust constraint on CO_2 fluxes. For example, the GOSAT-2 satellite will monitor atmospheric CO_2 with better accuracy relative to the original GOSAT satellite (**?**). This improvement in both the quality and overall quantity of CO_2 observations will enable more detailed estimates of CO_2 fluxes. In addition to GOSAT-2, the OCO-3 mission will observe CO_2 from the International Space Station at a different locations and times of day relative to OCO-2 (NASA Jet Propulsion Laboratory, 2017). This feature will provide a stronger constraint on spatial and temporal variations in CO_2 fluxes. ~~In addition, observation or~~ However, retrieval errors appear to be a key factor in our results and will likely be a challenge for these future missions. Work on the OCO-2 retrieval algorithm will inform these upcoming missions, so improvements to the OCO-2 retrievals will likely improve the data capabilities of future missions as well. Further improvements to the satellite retrieval and atmospheric transport modeling could enable OCO-2 and future missions to provide detailed CO_2 budgets for much finer regions.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank David Baker, David Crisp, Benjamin Poulter, and Eva Sinha for their help and feedback on the manuscript. The OCO-2 data are produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. OCO-2 data are publicly available online at https://co2.jpl.nasa.gov/#mission=OCO-2. The MODIS EVI and NDVI data products are available online at https://lpdaac.usgs.gov/ courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. GOME-2 SIF data are available at https://acd-ext.gsfc.nasa.gov/People/Joiner/my_gifs/GOME_F/GOME-F.htm courtesy of Joanna Joiner. MsTMIP data products are archived at the ORNL DAAC (http://daac.ornl.gov), and 3-hourly flux products are available at dx.doi.org/10.3334/ORNLDAAC/1315. This work is funded by the Carnegie Distinguished Postdoctoral Fellowship and NASA grant #NNX13AC48G. Funding for MsTMIP was provided through NASA ROSES Grant #NNX10AG01A. Data management support for preparing, documenting, and distributing model driver and output data was performed by the Modeling and Synthesis Thematic Data Center at Oak Ridge National Laboratory (ORNL; http://nacp.ornl.gov), with funding thorugh NASA ROSES Grant #NNH10AN681.

References

Alkhaled, A. A., Michalak, A. M., Kawa, S. R., Olsen, S. C., and Wang, J.-W.: A global evaluation of the regional spatial variability of column integrated CO_2 distributions, J. Geophys. Res. – Atmos., 113, doi:10.1029/2007JD009693, d20303, 2008.

ASCENDS Ad Hoc Science Definition Team: Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (ASCENDS) Mission: Science Mission Definition Study, Tech. rep., NASA, http://cce.nasa.gov/ascends_2015/ASCENDS_FinalDraft_4_27_15.pdf, 2015.

Baker, D. F., Bösch, H., Doney, S. C., O'Brien, D., and Schimel, D. S.: Carbon source/sink information provided by column CO_2 measurements from the Orbiting Carbon Observatory, Atmos. Chem. Phys., 10, 4145–4165, doi:10.5194/acp-10-4145-2010, 2010.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.

Basu, S., Krol, M., Butz, A., Clerbaux, C., Sawa, Y., Machida, T., Matsueda, H., Frankenberg, C., Hasekamp, O. P., and Aben, I.: The seasonal variation of the CO_2 flux over Tropical Asia estimated from GOSAT, CONTRAIL, and IASI, Geophys. Res. Lett., 41, 1809–1815, doi:10.1002/2013GL059105, 2014.

Chatterjee, A., Gierach, M. M., Sutton, A. J., Feely, R. A., Crisp, D., Eldering, A., Gunson, M. R., O'Dell, C. W., Stephens, B. B., and Schimel, D. S.: Influence of El Niño on atmospheric CO_2 over the tropical Pacific Ocean: Findings from NASA's OCO-2 mission, Science, 358, doi:10.1126/science.aam5776, 2017.

Chevallier, F., Engelen, R. J., and Peylin, P.: The contribution of AIRS data to the estimation of CO_2 sources and sinks, Geophys. Res. Lett., 32, doi:10.1029/2005GL024229, l23801, 2005.

Chevallier, F., Bron, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res. – Atmos., 112, doi:10.1029/2006JD007375, d09307, 2007.

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 41, 1065–1070, doi:10.1002/2013GL058772, 2013GL058772, 2014.

Crisp, D., Atlas, R., Breon, F.-M., Brown, L., Burrows, J., Ciais, P., Connor, B., Doney, S., Fung, I., Jacob, D., Miller, C., O'Brien, D., Pawson, S., Randerson, J., Rayner, P., Salawitch, R., Sander, S., Sen, B., Stephens, G., Tans, P., Toon, G., Wennberg, P., Wofsy, S., Yung, Y., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Adv. Space Res., 34, 700 – 709, doi:10.1016/j.asr.2003.08.062, 2004.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M., Sun, K., Taylor, T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmos. Meas. Tech., 10, 59–81, doi:10.5194/amt-10-59-2017, 2017.

Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO_2 from GOSAT XCO_2 data, Atmos. Chem. Phys., 14, 3703–3727, doi:10.5194/acp-14-3703-2014, 2014.

Eldering, A., Boland, S., Solish, B., Crisp, D., Kahn, P., and Gunson, M.: High precision atmospheric CO_2 measurements from space: The design and implementation of OCO-2, in: 2012 IEEE Aerospace Conference, pp. 1–10, doi:10.1109/AERO.2012.6187176, 2012.

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

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., Liu, J., Schwandner, F. M., Sun, Y., O'Dell, C. W., Frankenberg, C., Taylor, T., Fisher, B., Osterman, G. B., Wunch, D., Hakkarainen, J., Tamminen, J., and Weir, B.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, Science, 358, doi:10.1126/science.aam5745, 2017b.

European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL): Emission Database for Global Atmospheric Research (EDGAR), release EDGARv4.2 FT2010, http://edgar.jrc.ec.europa.eu, 2013.

Fang, Y. and Michalak, A. M.: Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers, Global Biogeochem. Cy., 29, 555–566, doi:10.1002/2014GB005034, 2014GB005034, 2015.

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

Fischer, M. L., Parazoo, N., Brophy, K., Cui, X., Jeong, S., Liu, J., Keeling, R., Taylor, T. E., Gurney, K., Oda, T., and Graven, H.: Simulating estimation of California fossil fuel and biosphere carbon dioxide exchanges combining in situ tower and satellite column observations, J. Geophys. Res. – Atmos., 122, 3653–3671, doi:10.1002/2016JD025617, 2016JD025617, 2017.

Fisher, J. B., Sikka, M., Huntzinger, D. N., Schwalm, C., and Liu, J.: Technical note: 3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange, Biogeosciences, 13, 4271–4277, doi:10.5194/bg-13-4271-2016, 2016.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res. – Biogeo., 118, 317–328, doi:10.1002/jgrg.20042, 2013.

Gourdji, S. M., Mueller, K. L., Schaefer, K., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, J. Geophys. Res. – Atmos., 113, doi:10.1029/2007JD009733, d21115, 2008.

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

Griffith, D. A.: Effective geographic sample size in the presence of spatial autocorrelation, Ann. Assoc. Am. Geogr., 95, 740–760, doi:10.1111/j.1467-8306.2005.00484.x, 2005.

Guerlet, S., Basu, S., Butz, A., Krol, M., Hahne, P., Houweling, S., Hasekamp, O. P., and Aben, I.: Reduced carbon uptake during the 2010 Northern Hemisphere summer from GOSAT, Geophys. Res. Lett., 40, 2378–2383, doi:10.1002/grl.50402, 2013.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., et al.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, doi:10.1038/415626a, 2002.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., and de la Rue du Can, S.: High resolution fossil fuel combustion CO_2 emission fluxes for the United States, Environ. Sci. Technol., 43, 5535–5541, doi:10.1021/es900806c, pMID: 19708393, 2009.

Hammerling, D. M., Michalak, A. M., and Kawa, S. R.: Mapping of CO_2 at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2, J. Geophys. Res. – Atmos., 117, doi:10.1029/2011JD017015, d06306, 2012.

Hammerling, D. M., Kawa, S. R., Schaefer, K., Doney, S., and Michalak, A. M.: Detectability of CO_2 flux signals by a space-based lidar mission, J. Geophys. Res.-Atmos., 120, 1794–1807, doi:10.1002/2014JD022483, 2014JD022483, 2015.

Heymann, J., Reuter, M., Buchwitz, M., Schneising, O., Bovensmann, H., Burrows, J. P., Massart, S., Kaiser, J. W., and Crisp, D.: CO_2 emission of Indonesian fires in 2015 estimated from satellite-derived atmospheric CO_2 concentrations, Geophys. Res. Lett., 44, 1537–1544, doi:10.1002/2016GL072042, 2016GL072042, 2017.

Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L., Ganshin, A., Hasekamp, O., Jones, D., Maksyutov, S., Marshall, J., Oda, T., O'Dell, C. W., Oshchepkov, S., Palmer, P. I., Peylin, P., Poussi, Z., Reum, F., Takagi, H., Yoshida, Y., and Zhuravlev, R.: An intercomparison of inverse models for estimating sources and sinks of CO_2 using GOSAT measurements, J. Geophys. Res. – Atmos., 120, 5253–5266, doi:10.1002/2014JD022962, 2014JD022962, 2015.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens. Environ., 83, 195 – 213, doi:10.1016/S0034-4257(02)00096-2, 2002.

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., Jacobson, A., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang, M., Ito, A., Lei, H., Lu, C., Mao, J., Peng, C. H., Peng, S., Poulter, B., Riccuito, D., Shi, X., Tian, H., Wang, W., Zeng, N., Zhao, F., and Zhu, Q.: The North American Carbon Program Multi-Scale Synthesis and Terrestrial Model Intercomparison Project – Part 1: Overview and experimental design, Geosci. Model Dev., 6, 2121–2133, doi:10.5194/gmd-6-2121-2013, 2013.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., 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, doi:10.5194/amt-6-2803-2013, 2013.

Jones, R. H.: Bayesian information criterion for longitudinal and clustered data, Stat. Med., 30, 3050–3056, doi:10.1002/sim.4323, 2011.

Kawa, S. R., Erickson, D. J., Pawson, S., and Zhu, Z.: Global CO_2 transport simulations using meteorological data from the NASA data assimilation system, J. Geophys. Res. – Atmos., 109, doi:10.1029/2004JD004554, d18312, 2004.

Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S.-J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Patra, P. K., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Takigawa, M., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO_2: Experimental overview and diurnal cycle results for 2002, Global Biogeochem. Cy., 22, doi:10.1029/2007GB003050, gB3009, 2008.

Liang, A., Gong, W., Han, G., and Xiang, C.: Comparison of Satellite-Observed XCO_2 from GOSAT, OCO-2, and Ground-Based TCCON, Remote Sensing, 9, 2017.

Liu, J., Fung, I., Kalnay, E., and Kang, J.-S.: CO_2 transport uncertainties from the uncertainties in meteorological fields, Geophys. Res. Lett., 38, doi:10.1029/2011GL047213, l12808, 2011.

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, doi:10.1126/science.aam5690, 2017.

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_2 flux estimates for 2009–2010 based on GOSAT and ground-based CO_2 observations, Atmos. Chem. Phys., 13, 9351–9373, doi:10.5194/acp-13-9351-2013, 2013.

Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I., and Liu, J.: Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors, Atmos. Chem. Phys., 15, 2903–2914, doi:10.5194/acp-15-2903-2015, 2015.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cy., 24, doi:10.1029/2009GB003642, gB3023, 2010.

NASA Jet Propulsion Laboratory: Orbiting Carbon Observatory 3, http://www.jpl.nasa.gov/missions/orbiting-carbon-observatory-3-oco-3/, 2017.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying CO_2 Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045–10,053, doi:10.1002/2017GL074702, 2017GL074702, 2017.

National Research Council: Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, The National Academies Press, Washington, DC, doi:10.17226/12883, http://www.nap.edu/catalog/12883/verifying-greenhouse-gas-emissions-methods-to-support-international-climate-agreements, 2010.

Olivier, J., Janssens-Maenhout, G., Munteam, M., and Peters, J.: Trends in global CO_2 emissions; 2014 Report, PBL Netherlands Environmental Assessment Agency; European Commission, Joint Research Centre, Ispra, Italy, 2014.

Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO_2 and implications for carbon cycle research, Journal of Geophysical Research: Atmospheres, 109, doi:10.1029/2003JD003968, d02301, 2004.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial ecoregions of the world: A new map of life on Earth, BioScience, 51, 933–938, doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

Parazoo, N. C., Bowman, K., Frankenberg, C., Lee, J.-E., Fisher, J. B., Worden, J., Jones, D. B. A., Berry, J., Collatz, G. J., Baker, I. T., Jung, M., Liu, J., Osterman, G., O'Dell, C., Sparks, A., Butz, A., Guerlet, S., Yoshida, Y., Chen, H., and Gerbig, C.: Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of XCO_2 and chlorophyll fluorescence from GOSAT, Geophys. Res. Lett., 40, 2829–2833, doi:10.1002/grl.50452, 2013.

Patra, P. K., Crisp, D., Kaiser, J. W., Wunch, D., Saeki, T., Ichii, K., Sekiya, T., Wennberg, P. O., Feist, D. G., Pollard, D. F., Griffith, D. W. T., Velazco, V. A., De Maziere, M., Sha, M. K., Roehl, C., Chatterjee, A., and Ishijima, K.: The Orbiting Carbon Observatory (OCO-2) tracks 2–3 peta-gram increase in carbon release to the atmosphere during the 2014–2016 El Niño, Scientific Reports, 7, 13 567, doi:10.1038/s41598-017-13459-0, 2017.

Ramsey, F. and Schafer, D.: The Statistical Sleuth: A Course in Methods of Data Analysis, Cengage Learning, Boston, MA, 2012.

Rayner, P. J. and O'Brien, D. M.: The utility of remotely sensed CO_2 concentration data in surface source inversions, Geophysical Research Letters, 28, 175–178, http://dx.doi.org/10.1029/2000GL011912, 2001.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Saeki, T., Maksyutov, S., Saito, M., Valsala, V., Oda, T., Andres, R. J., Belikov, D., Tans, P., Dlugokencky, E., Yoshida, Y., Morino, I., Uchino, O., and Yokota, T.: Inverse modeling of CO_2 fluxes using GOSAT data and multi-year ground-based observations, SOLA, 9, 45–50, doi:10.2151/sola.2013-011, 2013.

Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E., and Worthy, D. E. J.: A regional high-resolution carbon flux inversion of North America for 2004, Biogeosciences, 7, 1625–1644, doi:10.5194/bg-7-1625-2010, 2010.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, http://www.jstor.org/stable/2958889, 1978.

Shiga, Y. P., Michalak, A. M., Randolph Kawa, S., and Engelen, R. J.: In-situ CO2 monitoring network evaluation and design: A criterion based on atmospheric CO2 variability, J. Geophys. Res. – Atmos., 118, 2007–2018, doi:10.1002/jgrd.50168, 2013.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014GL059684, 2014.

Takagi, H., Saeki, T., Oda, T., Saito, M., Valsala, V., Belikov, D., Saito, R., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., Yokota, T., and Maksyutov, S.: On the benefit of GOSAT observations to the estimation of regional CO_2 fluxes, SOLA, 7, 161–164, doi:10.2151/sola.2011-041, 2011.

Takahashi, T., Sutherland, S., and Kozyr, A.: Global Ocean Surface Water Partial Pressure of CO_2 Database: Measurements Performed During 1957-2015 (Version 2015)., Tech. Rep. ORNL/CDIAC-160, NDP-088(V2015), Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennesee, doi:10.3334/CDIAC/OTG.NDP088(V2015), 2016.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11 707–11 735, doi:10.5194/acp-10-11707-2010, 2010.

Worden, J. R., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., O'Dell, C., and Bowman, K.: Evaluation and attribution of OCO-2 XCO_2 uncertainties, Atmospheric Measurement Techniques, 10, 2759–2771, doi:10.5194/amt-10-2759-2017, 2017.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., Macatangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T., and Wofsy, S. C.: A method for evaluating bias in global measurements of CO_2 total columns from space, Atmos. Chem. Phys., 11, 12 317–12 337, doi:10.5194/acp-11-12317-2011, 2011.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON, Atmospheric Measurement Techniques, 10, 2209–2238, doi:10.5194/amt-10-2209-2017, 2017a.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) $X_{\mathrm{CO_2}}$ measurements with TCCON, Atmos. Meas. Tech., 10, 2209–2238, doi:10.5194/amt-10-2209-2017, 2017b.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., and Maksyutov, S.: Global Concentrations of CO_2 and CH_4 Retrieved from GOSAT: First Preliminary Results, SOLA, 5, 160–163, doi:10.2151/sola.2009-041, 2009.

Zucchini, W.: An Introduction to Model Selection, J. Math. Psychol., 44, 41 – 61, doi:10.1006/jmps.1999.1276, 2000.

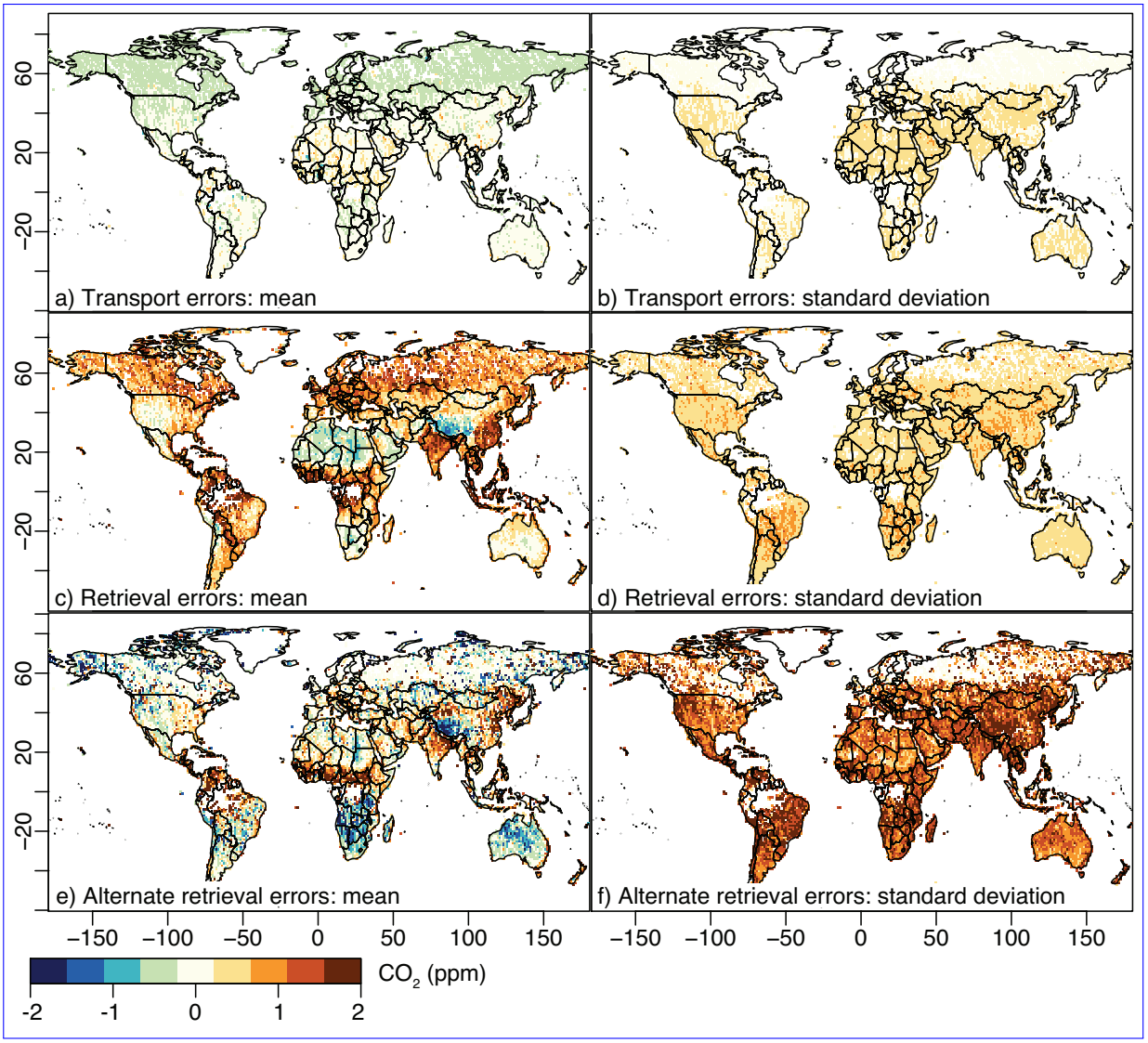

Figure 1. This figure displays the simulated synthetic data errors (i.e., simulated ϵ): the simulated atmospheric transport errors (a-b), the simulated observation/retrieval errors (c-d), and a second, alternative set of observation/retrieval errors (e-f). The left-hand panels display the mean of the errors within each PCTM grid box across the entire 2014–2015 study period, an indication of the correlation or covariance among the errors. By contrast, the right-hand panels display the standard deviation of the errors or residuals within each PCTM grid box.

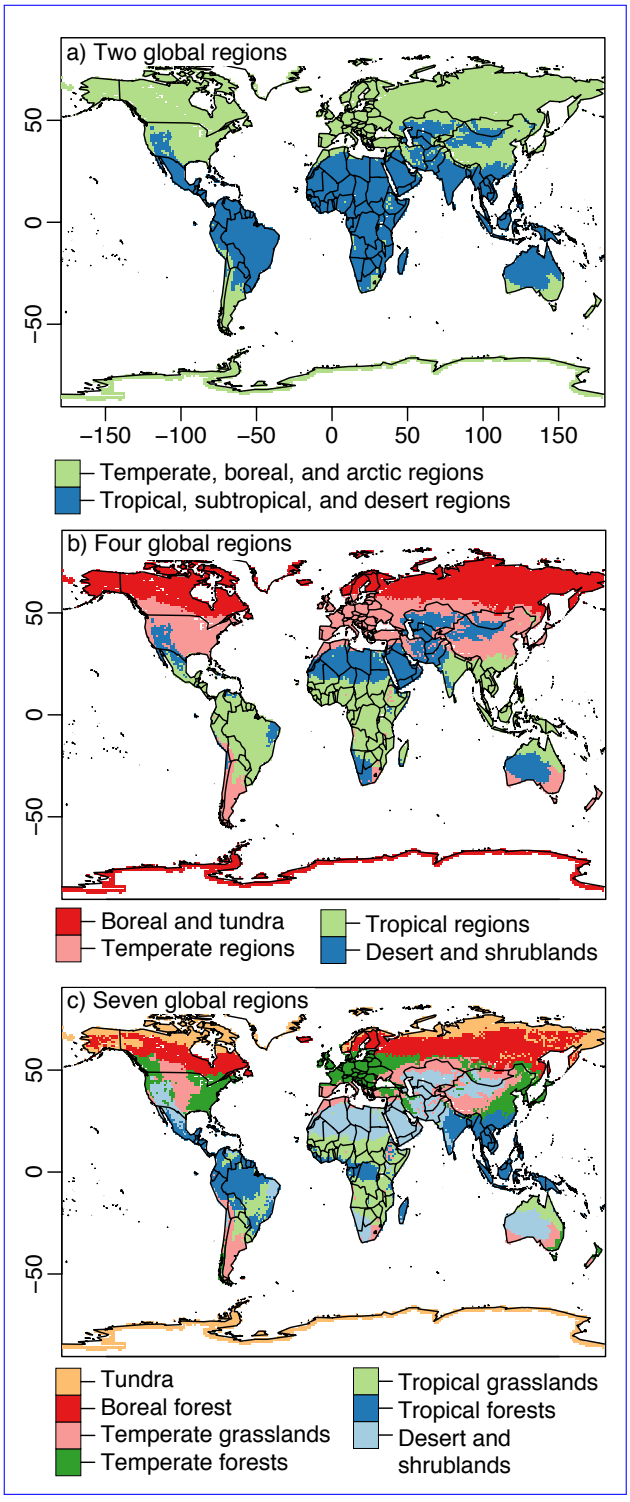

Figure 2. The two hemispheric regions (a), four continental regions (b), and seven ecoregions (c) used in this study. The ecoregions are based on a world biome map by Olson et al. (2001). The two and four region maps are amalgamated versions of the ecoregions.

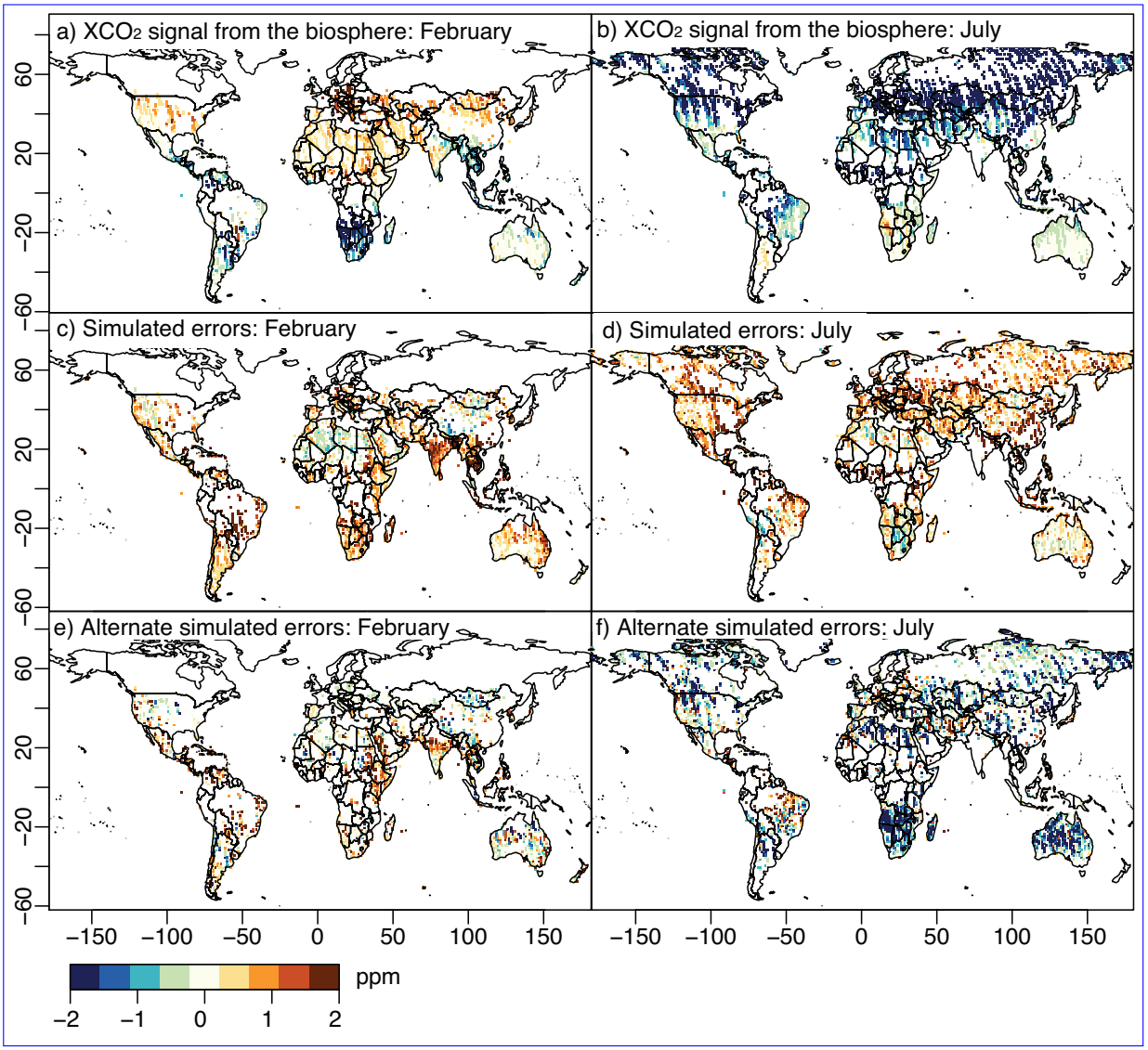

Figure 3. This figure compares the total column CO_2 or XCO_2 signal from biospheric fluxes against simulated model and observation errors. Panels a and b display the monthlong mean XCO_2 signal from biospheric CO_2 fluxes for February and July, respectively. We estimate this signal using the SiBCASA flux model and PCTM. Also note that the XCO_2 signal for February and July includes CO_2 fluxes from the months of February and July, respectively, and no fluxes from previous or subsequent months. Panels c and d represent the sum of both simulated observation and atmospheric transport errors (monthlong mean, Sect. 2.2). Panels e and f show the sum of these errors using an alternate estimate for the retrieval errors.

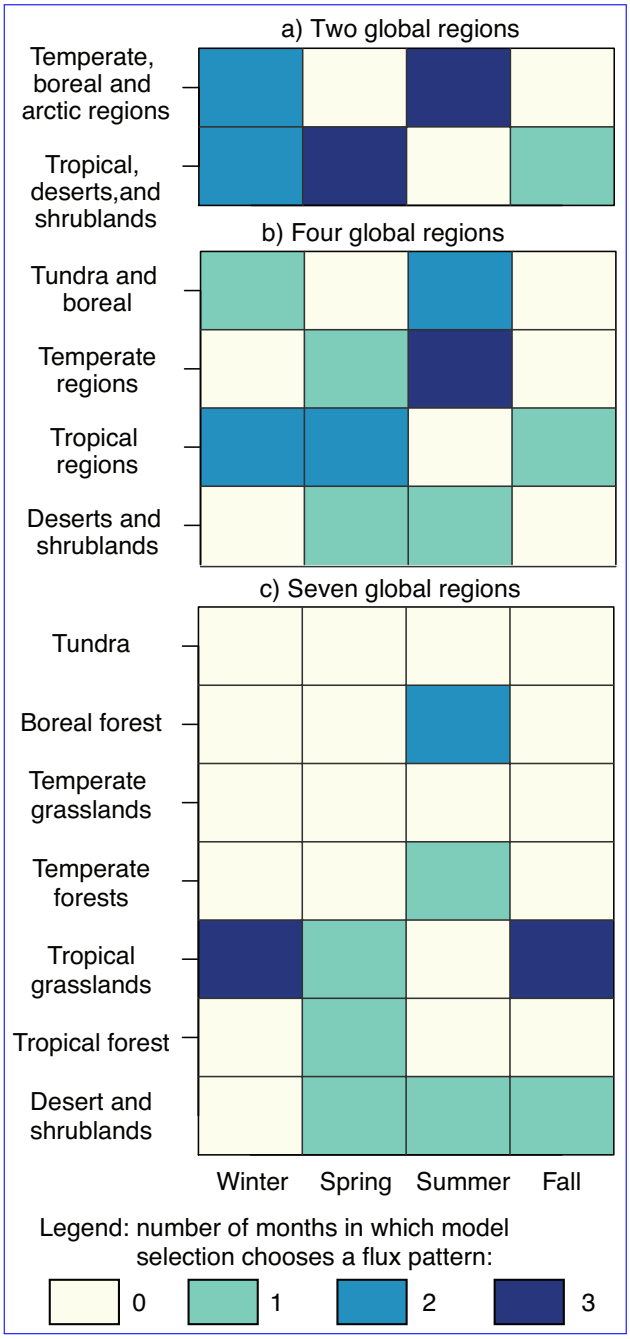

Figure 4. The results of the model selection experiments using real data. ~~We select~~ The different colors indicate the number of months in which at least one flux pattern is selected for a given region, and dark colors suggest excellent detectability while light colors suggest limited detection abilities. Flux patterns are selected for a greater fraction of regions/months in the two region case (a) than in the four or seven region cases (b and c).~~The different colors indicate the number of months in which model selection chooses a flux pattern for a given region.~~

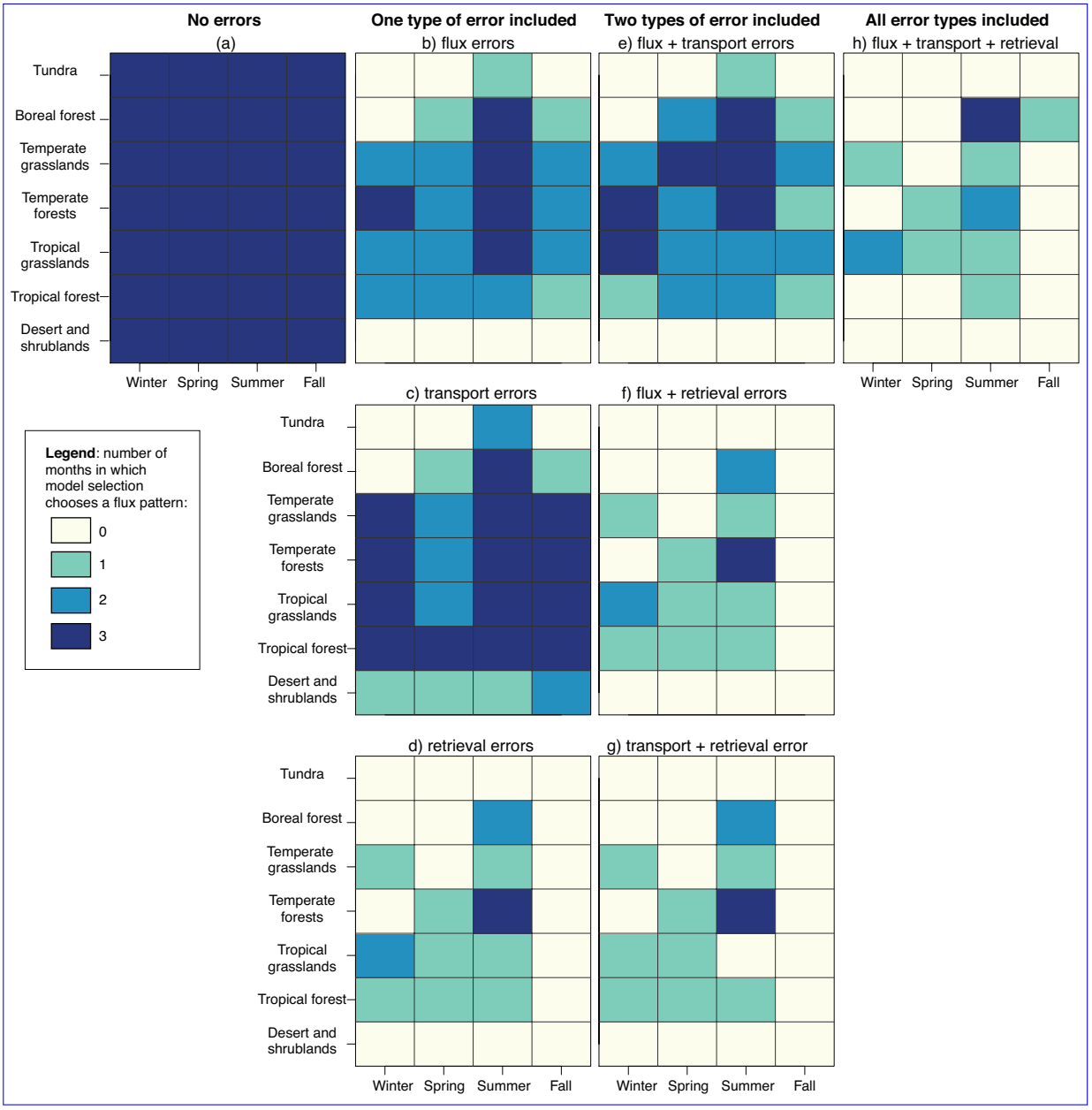

Figure 5. Model selection results for the synthetic data experiments. The first column (a) shows an experiment with no errors in the synthetic observations. The results of that experiment are ideal, and ~~we choose a~~ at least one flux pattern is selected in every month and every region using model selection. Subsequent panels (b-h) show the results with one, two, and three types of errors included. Fewer regions and seasons are selected in these experiments. The plot signals that ~~observation or~~ retrieval errors likely play a key role.

S1 Additional detail on study methodology

S1.1 Details on the CO_2 initial condition and model spin-up

We spin up CO_2 mixing ratios in the PCTM model using outputs from NOAA's CarbonTracker product (Peters et al., 2007). CarbonTracker produces CO_2 fluxes and atmospheric mole fractions that are optimized to match available in situ CO_2 observations. Our goal is to spin up CO_2 mixing ratios within PCTM in a fashion that is both consistent with CarbonTracker and with the PCTM model grid. To this end, we initialize CO_2 simulations on 1 Jan., 2009. We average estimated global CO_2 mole fractions from CarbonTracker by hemisphere and by model vertical level. We then use these averages as the CO_2 initial condition in PCTM for 1 Jan., 2009. Subsequently, we run PCTM forward using CarbonTracker fluxes until 1 Sept., 2014 when the model selection simulations begin. We continue these PCTM simulations past 1 Sept., 2014 using a surface flux of zero. These simulations become the spin-up modeled CO_2, and we subtract these spin-up mole fractions from the OCO-2 retrievals ($\boldsymbol{z}$ in Eq. S1). The resulting vector ($\boldsymbol{z}$) represents the change in CO_2 mole fractions due to fluxes that occurred after 1 Sept., 2014.

S1.2 Additional detail on the CO_2 flux patterns used in model selection

This section provides additional detail on the terrestrial biosphere models (TBMs) and vegetation indices that are used in the model selection experiments. These TBMs and vegetation indices are used as the input fluxes in the PCTM model. We generate modeled XCO_2 total columns using these PCTM outputs, and the modeled XCO_2 total columns become the predictor variables in the model selection experiments (i.e., the columns of the $\mathbf{X}$ matrix, Eq. S1). Note that the multiple regression will scale the magnitude of each column of $\mathbf{X}$ in each region and each month to best match the observations (Eq. S1). As a result of this setup, the overall magnitude of each TBM and of each vegetation index does not affect the model selection results. Rather, this study utilizes the spatial and temporal patterns in the TBMs and vegetation indices.

We include four TBMs from the recent MsTMIP project (Huntzinger et al., 2013). The selected TBMs have very different space-time patterns and therefore sample a wide range of plausible flux patterns. These TBMs include the Dynamic Land Ecosystem Model (DLEM; e.g., Tian et al., 2011), the Lund-Potsdam-Jena Model Wald Schnee und Landschaft version (LPJ; e.g., Sitch et al., 2003), the Global Terrestrial Ecosystem Carbon Model (GTEC; e.g., King et al., 1997), and the Simple Biosphere Model with the Carnegie-Ames-Stanford Approach (SIBCASA; e.g., Schaefer et al., 2008). The original MsTMIP model outputs have a spatial resolution of 0.5° latitude by 0.5° longitude and a 3-hourly temporal resolution. We regrid the fluxes to the PCTM model grid (1° latitude by 1.25° longitude) and input the fluxes into PCTM at the original 3-hourly resolution. Note that the gridded, 3-hourly MsTMIP flux estimates are available for years 2004–2010. Few TBMs have readily-available flux estimates for the 2014–2015 time period of this study, including the TBMs in the MsTMIP study. Instead, we use a multi-year average of the MsTMIP fluxes as inputs in the PCTM model. We average these 7 years within each separate grid box and for each separate 3-hourly time period to produce this multi-year average for each MsTMIP flux model.

In addition to these TBMs, we also utilize several vegetation indices (EVI, NDVI, and SIF). Numerous studies indicate that biospheric CO_2 fluxes correlate with these vegetation indices – with EVI (e.g., Sims et al., 2008; Wu et al., 2011), NDVI (e.g., Cihlar et al., 1992; Wylie et al., 2003), and GOME-2 SIF (e.g., Guanter et al., 2014; Yang et al., 2015). These indices are therefore good candidate flux patterns to use within the model selection experiments.

We use EVI and NDVI from the MODIS Aqua product MYD13C1 (Didan, 2015a) and the MODIS Terra product MOD13C1 (Didan, 2015b). These products are collectively available at 8-day intervals. The individual Aqua and Terra products are each available at 16-day intervals. However, the two products are staggered, so Aqua and Terra can be combined to produce EVI and NDVI estimates every 8 days. These products have a 0.05° latitude by 0.05° longitude, and we regrid them to the PCTM model grid (1° latitude by 1.25° longitude). Both of these products are available for 2014 and 2015, the time period of this study.

We use level 2 SIF retrievals from GOME-2 (Global Ozone Monitoring Experiment-2) (Joiner, 2014). We convert the level 2 retrievals to a gridded SIF product using a block kriging method described by Tadić et al. (2017). This gridded product has a daily temporal resolution and the same spatial resolution as PCTM. We use this product as an input 'flux' into the PCTM model and incorporate the PCTM outputs as candidate variables in the $\mathbf{X}$ matrix.

S1.3 Additional detail on model selection implementation

This sub-section describes the regression and model selection in greater detail. The regression used in this paper will quantitative link OCO-2 XCO_2 observations with atmospheric model outputs:

$$\mathbf{Y} = h(\mathbf{X}) \tag{S1}$$

$$\boldsymbol{z} = \mathbf{Y}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{S2}$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V}) \tag{S3}$$

These equations are an expanded form of the regression equations present in Sect. 2.3. The vector $\boldsymbol{z}$ (dimensions $n \times 1$) represents the XCO_2 observations minus the model initial condition or spin-up (Sect. S1.1). The variable $\mathbf{X}$ (dimensions $m \times p$) is a matrix of p different flux models, and each column of $\mathbf{X}$ is a different flux model for a different region and month. The function $h()$ is an atmospheric model that transports the fluxes to the times and locations of the OCO-2 retrievals, and the resulting matrix $\mathbf{Y}$ has dimensions $n \times p$. Furthermore, the variable $\boldsymbol{\epsilon}$ is a $n \times 1$ vector of residuals. These residuals are assumed to follow a multivariate normal distribution with a mean of zero, a variance of σ^2, and a covariance structure given by $\mathbf{V}$ (dimensions $n \times n$). The vector of coefficients ($\boldsymbol{\beta}$, dimensions $p \times 1$) are estimated as part of the regression.

In this study, we choose a set of variables for $\mathbf{X}$ using model selection based on the Bayesian Information Criterion (BIC) (Schwarz, 1978). We calculate a BIC score for many different candidate models. Each candidate model has a different set of columns ($\mathbf{X}$) – different combinations of flux models in different geographic regions and in different months.

The best model has the lowest BIC score:

$$BIC = L + p\ln(n^*) \tag{S4}$$

where L is the log likelihood of a particular candidate model ($\mathbf{X}$). The log-likelihood has the following form:

$$L = n^*\ln(\sigma^2) + \frac{n^*}{n}RSS \tag{S5}$$

$$RSS = \frac{1}{\sigma^2}\boldsymbol{z}^T\boldsymbol{z} - \frac{1}{\sigma^2}\boldsymbol{z}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\boldsymbol{z} \tag{S6}$$

where RSS is the residual sum of squares and σ^2 is defined above in Eq. S3.

Both the BIC and log-likelihood equations (Eq. S4 and S6) incorporate n^*, the effective number of independent observations. Jones (2011) discusses this concept in the context of the BIC. Just because the satellite provides n observations does not mean there are n independent pieces of information. Accordingly, n^* ensures that the model selection framework accurately assesses the amount of independent information in the observations. It accounts for the fact that there are often spatial and temporally coherent errors in the satellite observations or in the transport model. If all of the observations were independent (i.e., if $\mathbf{V}$ were diagonal), then n^* would equal n. However, we de-weight both components of Eq. S4 as the covariances in $\mathbf{V}$ increase.

We could calculate n^* directly using $\mathbf{V}^{-1}$ (Jones, 2011). In fact, several CO_2 model selection studies incorporate $\mathbf{V}^{-1}$ directly into the equation for RSS (e.g., Mueller et al., 2010; Gourdji et al., 2012; Shiga e
We use 5,079,165 observations (n) in this study, so $\mathbf{V}$ has 5.08×10^6 rows and columns. As a result, the inverse of $\mathbf{V}$ is computationally intractable. We instead estimate n^* using an approach based on Griffith (2005), an approach that does not require computing $\mathbf{V}^{-1}$ directly:

$$n^* = \frac{n}{1 + \left(\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} V_{i,j} / n\right)} \tag{S7}$$

This equation calculates n^* using individual elements of $\mathbf{V}$ and does not require inverting the full matrix; it is therefore far more computationally tractable. Subsequent paragraphs describe how we estimate the elements of $\mathbf{V}$.

The remainder of this section discusses how we characterize the variances (σ^2) and covariance structure ($\mathbf{V}$) of the ~~model-data~~ residuals ($\boldsymbol{\epsilon}$) (Eq. S3). An estimate of the variance is required to calculate the residual sum of squares (RSS, Eq. S6), and an estimate of the covariance structure is necessary to calculate the effective number of independent observations (n^*, Eq. S7).

We first describe ~~our approach to~~ the method for characterizing the covariance structure ($\mathbf{V}$). We estimate ~~this structure by constructing empirical variograms and fitting spherical variogram models (for an overview of variograms, refer to Kitanidis, 1997). We use a spherical model because it tapers off to zero and is therefore faster to compute with large datasets. We fit these variagrams locally~~ the individual elements, $V_{i,j}$, in the vicinity of each ~~residual~~ observation i ~~, where i denotes an individual element of~~ by fitting a local variogram model on the model-data residuals ($\boldsymbol{\epsilon}$). The covariance structure likely differs in different locations and at different times (i.e., is non-stationary)~~. Several~~ , and several existing studies fit variograms locally to account for this non-stationary structure ~~(e.g., Hammerling et al., 2012; Tadić et al., 2017), and we~~ (e.g., Alkhaled et al., 2008; Hammerling et al., 2012; Tadić et al., 2017). We use a similar approach here. Specifically, we estimate this structure by constructing empirical variograms and fitting spherical variogram models (for an overview of variograms, refer to Kitanidis, 1997). We use a spherical model because it tapers off to zero and is therefore faster to compute with large datasets.

For each i, we create a separate spatial experimental variogram and a temporal experimental variogram. We use all residuals that lie within 3000 kilometers of i and were collected at similar times (within 1 day) to construct the spatial variogram, and we use residuals within 75 days of i and a similar location (within 250 km) to construct the spatial variogram. We choose these spatial and temporal distances because they are larger than transport or retrieval errors that might covary across an entire ecoregion. We then fit spherical models to the spatial and temporal variograms, respectively. In our setup, the elements of $V_{i,j}$ equal the spherical

covariance model multiplied by the temporal covariance model:

$$V_{i,j} = \begin{cases} (1 - \frac{3}{2}\frac{d_{i,j}}{\alpha_{d,i}} + \frac{1}{2}\frac{d_{i,j}^3}{\alpha_{d,i}^3})(1 - \frac{3}{2}\frac{t_{i,j}}{\alpha_{t,i}} + \frac{1}{2}\frac{t_{i,j}^3}{\alpha_{t,i}^3}) & \text{if } d \leq \alpha_{d,i} \text{ and } t \leq \alpha_{t,i} \\ 0 & \text{if } d > \alpha_{d,i} \text{ or } t > \alpha_{t,i} \end{cases}$$

where $d_{i,j}$ and $t_{i,j}$ are the distance and time, respectively, between points i and j. The variables $\alpha_{d,i}$ and $\alpha_{t,i}$ are the decorrelation length and time parameters estimated from the spatial and temporal emperical variograms, respectively. Note that several existing top-down CO_2 studies use a covariance model with multiplied spatial and temporal components (e.g., Mueller et al., 2008; Gourdji et al., 2012), though other recent studies use a more advanced approach (Tadić et al., 2017).

We then use this estimate for $V_{i,j}$ to estimate n^* (Eq. S7). The denominator of Eq. S7 sums over each element i. For each element i in the summation, we use covariance parameters estimated for that element i. Note that Eq. S7 presents a computational challenge. In this setup, $i = 1...n$ where $n = 5.08 \times 10^6$. It would be computationally prohibitive to estimate 5.08×10^6 local variogram models. Instead, we randomly choose 1000 elements of i and estimate a variogram model for each of those elements. As a result, the i summation term Eq. S7 only sums over 1000 elements. More precisely, Eq. S7 becomes

$$n^* = \frac{n}{1 + \frac{\sum_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} V_{i,j}}{1000}} \tag{S8}$$

where i is a set of 1000 randomly chosen numbers between 1 and n.

The model selection equations (i.e., Eq. S6) also require an estimate for σ^2. We estimate a single value for σ^2 using all n residuals:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} \epsilon_i^2 \tag{S9}$$

This implementation of model selection is iterative. We start by estimating the covariance parameters with all candidate variables included in $\mathbf{X}$. We use these covariance parameters to estimate n^* and subsequently run model selection. We run model selection with a heuristic branch and bound algorithm described by Yadav et al. (2013). This algorithm dramatically reduces the computing time of the model selection step. We then re-estimate the covariance parameters using the chosen columns of $\mathbf{X}$. The covariance parameters usually change slightly with the new, updated $\mathbf{X}$ matrix. We alternate between the covariance estimation and model selection until both the covariance parameters and columns selected for $\mathbf{X}$ do not change from one iteration to the next. The estimated covariance parameters and model selection results typically converge on a stable answer within two to three iterations.

S1.4 Additional detail on the simulated, synthetic data errors

This section provides additional information on the simulated atmospheric transport errors and simulated retrieval errors used in the synthetic data experiments (Sect. 2.2 and 2.5).

We use simulated transport errors from an ensemble of meteorology realizations in Miller et al. (2015). That study follows an approach developed by Liu et al. (2011). Both studies simulate global meteorology using the Community Atmosphere Model (CAM) in weather forecasting mode. The studies also include CO_2 as a passive tracer in the model. Miller et al. (2015) and Liu et al. (2011) then run an ensemble of 64 parallel simulations to estimate the effects of atmospheric transport uncertainties on modeled CO_2. At each time step of the simulations,

they assimilate the mean of the 64-member ensemble to match meteorological observations using a local ensemble Kalman filter (LETKF) (e.g., Hunt et al., 2007). Miller et al. (2015) also adjust the ensemble variance to be consistent with the meteorology model-data residuals using an approach known as adaptive covariance inflation (e.g., Miyoshi, 2011). In the present study, we randomly choose one of the ensemble members. We use the difference in modeled CO_2 mixing ratios between the chosen ensemble member and the ensemble mean as our simulated atmospheric transport error (Fig. 1). We then interpolate these estimated transport errors from the CAM model grid (1.9° latitude by 2.5° longitude resolution) to the locations and times of the GOSAT observations.

We use two different approaches to simulate satellite retrieval errors. We employ the first approach in the synthetic data simulations in the main text. We use the second set of errors here in the Supplement (Sect. S2.2) as a ~~consistency check on our~~ robustness or sensitivity check on the synthetic data simulations. In the first approach, we model XCO_2 using PCTM and the SiBCASA model. We then regress the model-data residuals on the retrieval parameters included in the OCO-2 lite data file. These parameters include retrieved surface pressure; the H_2O ratio; temperature at 700 hPa; wind speed; albedo; aerosol optical depth; the log of dust, water, and salt aerosols; the land fraction in the OCO-2 footprint; surface altitude; the satellite operation mode (e.g., nadir mode, target mode, etc.); footprint bias; and the change in CO_2 vertical gradient between the surface and retrieval model level 13.

We use this regression to estimate the portion of the residuals that map on to the retrieval parameters. That result is used as the estimated retrieval error (Fig. 1):

$$\boldsymbol{\epsilon} \;=\; \mathbf{1}\beta_{0,r} + \mathbf{X_r}\boldsymbol{\beta_r} + \zeta \tag{S10}$$

where $\boldsymbol{\epsilon}$ are the model-data residuals from Eq. S1, $\beta_{0,r}$ is the intercept term in the regression, $\mathbf{X_r}$ is the matrix of retrieval parameters, $\boldsymbol{\beta_r}$ are the estimated coefficients, and ζ is the portion the residuals ($\boldsymbol{\epsilon}$) not described any other terms in the equation. We use $\mathbf{X_r}\boldsymbol{\beta_r}$ as our estimate of the retrieval errors.

This approach is one way to estimate the portion of the residuals that maps onto retrieval parameters. Errors in the model output due to atmospheric transport or due to biospheric fluxes in $\mathbf{X}$ are unlikely to map onto parameters like aerosol optical depth. Rather, errors that map onto aerosol optical depth may more likely reflect issues in the satellite retrievals. There is always a possibility that residuals caused by inaccurate fluxes could have patterns similar to some of the retrieval parameters. For example, errors in modeled XCO_2 due to inaccurate biospheric flux patterns might correlate weakly or modestly with surface albedo, and errors in modeled XCO_2 due to errors in CO_2 fossil fuel emissions could correlate weakly or modestly with aerosol optical depth.

We generate an alternate set of retrieval errors using a different approach as a consistency check. This approach, described in Sect. 2.2, assigns a non-zero retrieval error if and only if four different flux estimates (input into PCTM) unanimously disagree with the OCO-2 observations. There could be error in the retrievals if all of the biospheric models are in good agreement – if all of the model outputs disagree with the retrievals in the same direction. In that case, either all four flux models are incorrect, there is a consistent transport bias, or there is an error in the satellite retrievals. The four flux models chosen in this study have disparate spatial and temporal patterns, so the first option appears unlikely. The second option (transport errors) could play a role, but transport errors at different vertical model levels often have different signs or magnitudes and can cancel out across the total column (e.g., Miller et al., 2015). Hence, the last option (retrieval errors) may be most likely when all four sets of model outputs consistently disagree with the XCO_2 retrievals. We use this alternative approach for generating retrieval errors solely as a consistency check on our primary synthetic data results.

S2 Sensitivity of the results to methodological choices

S2.1 Real data results that include glint data

We re-run model selection including all retrieval modes in the observation vector ($\boldsymbol{z}$) – glint, nadir, and target. The results in the main manuscript (Fig. 4), by contrast, exclude glint mode retrievals due to potential biases in these retrievals.

The model selection results that include glint retrievals are shown in Fig. S1~~. The~~, and the results look similar to those without glint retrievals (Fig. 4). Several factors may explain the similarity between these two results. The glint retrievals observe CO_2 mixing ratios in continental outflow but are not likely as sensitive to terrestrial flux patterns as nadir data over land. The model–data residuals over the ocean (i.e., associated with glint mode) are correlated over longer spatial and temporal scales relative to residuals over land. As a result, the glint mode retrievals add a limited amount of new, independent information on terrestrial fluxes in the context of the model selection experiments.

S2.2 Synthetic data results with an alternative set of simulated retrieval errors

Figure S2 displays the results of the synthetic data experiment using an alternate estimate for the retrieval errors~~. The results using this alternate error estimate~~, and the results are consistent with those in the main text (Fig. 5). This alternate set of simulated errors displays different characteristics from the errors used in Fig. 5, yet the model selection results are similar. These alternative simulated errors have a larger standard deviation but a smaller mean bias and smaller covariances (Fig. 1). Similarly, n^* is closer to the real data experiments ($n^* = 3700$). The larger standard deviation or variance increases the impact of these errors on the model selection results. However, smaller biases or covariances decrease their effect on the results. These two effects largely offset each other, and we therefore obtain similar results using both sets of simulated retrieval errors.

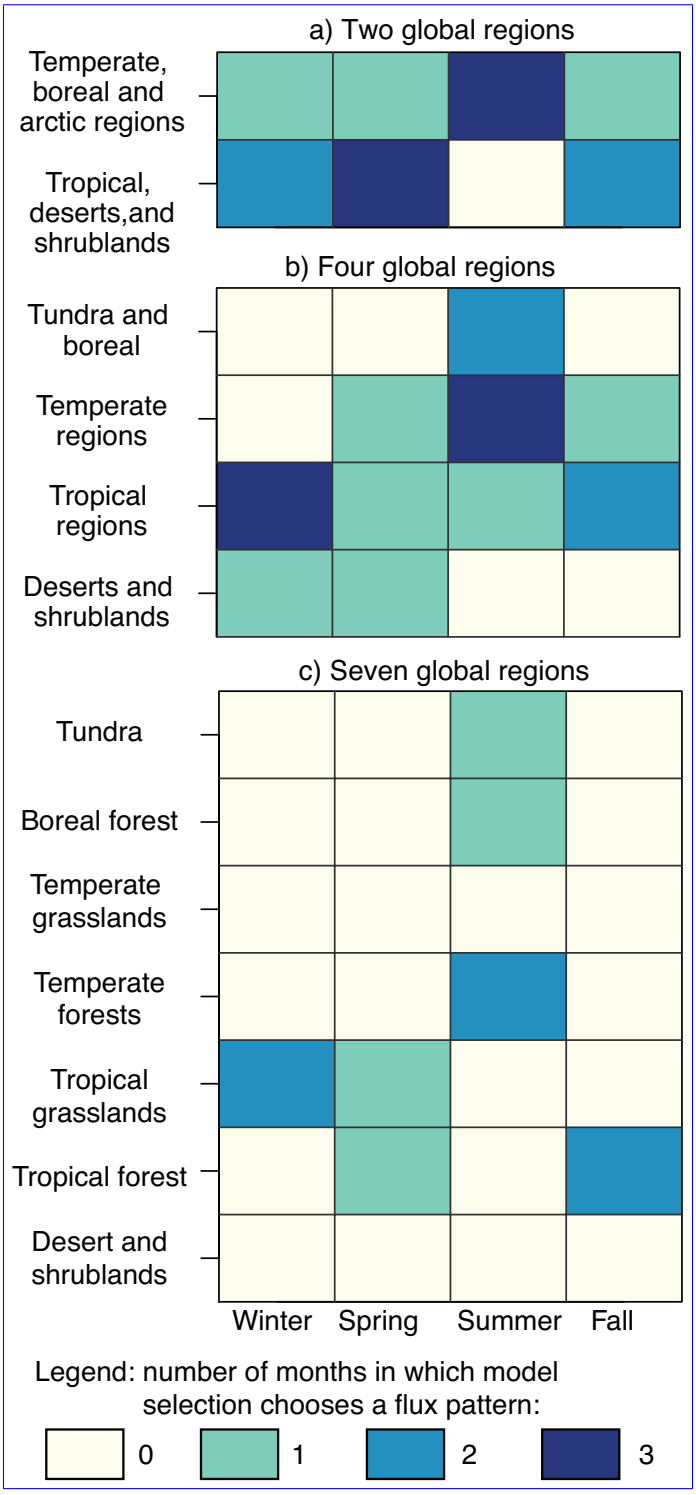

Figure S1: Results of the real data model selection experiment using all good quality OCO-2 retrievals (including glint mode retrievals). These model selection results are similar to those in the main manuscript that exclude glint mode data (Fig. 4).

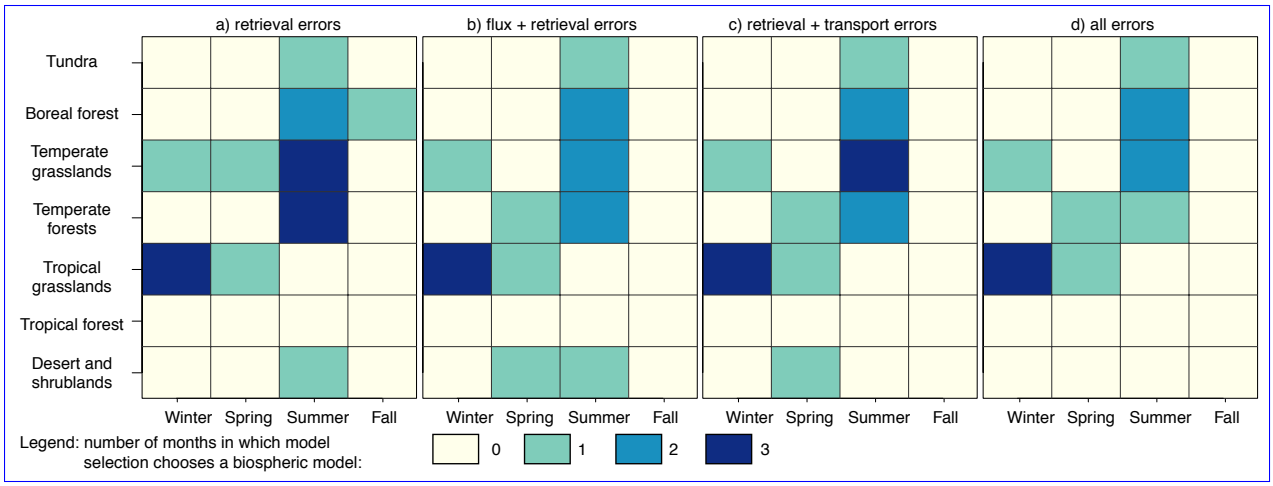

Figure S2: Results of the synthetic data case study using an alternative estimate for the retrieval errors (Sect. 2.2). These results are similar to those in Fig. 5, providing a consistency check on the synthetic data simulations.

References

Alkhaled, A. A., Michalak, A. M., Kawa, S. R., Olsen, S. C., and Wang, J.-W.: A global evaluation of the regional spatial variability of column integrated CO_2 distributions, J. Geophys. Res. – Atmos., 113, doi:10.1029/2007JD009693, d20303, 2008.

Cihlar, J., Caramori, P. H., Scnuepp, P. H., Desjardins, R. L., and MacPherson, J. I.: Relationship between satellite-derived vegetation indices and aircraft-based CO_2 measurements, J. Geophys. Res. – Atmos., 97, 18 515–18 521, doi:10.1029/92JD00655, 1992.

Didan, K.: MYD13C1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 0.05 Deg CMG V006, Tech. rep., NASA EOSDIS Land Processes DAAC, doi:10.5067/MODIS/MYD13C1.006, 2015a.

Didan, K.: MOD13C1 MODIS/Terra Vegetation Indices 16-Day L3 Global 0.05 Deg CMG V006, Tech. rep., NASA EOSDIS Land Processes DAAC, doi:10.5067/MODIS/MOD13C1.006, 2015b.

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

Griffith, D. A.: Effective geographic sample size in the presence of spatial autocorrelation, Ann. Assoc. Am. Geogr., 95, 740–760, doi:10.1111/j.1467-8306.2005.00484.x, 2005.

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J. M., and Griffis, T. J.: Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, P. Natl. Acad. Scie. USA, 111, E1327–E1333, doi:10.1073/pnas.1320008111, 2014.

Hammerling, D. M., Michalak, A. M., and Kawa, S. R.: Mapping of CO_2 at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2, J. Geophys. Res. – Atmos., 117, doi:10.1029/2011JD017015, d06306, 2012.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, doi:10.1016/j.physd.2006.11.008, 2007.

Huntzinger, D. N., Schwalm, C., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., Jacobson, A., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang, M., Ito, A., Lei, H., Lu, C., Mao, J., Peng, C. H., Peng, S., Poulter, B., Riccuito, D., Shi, X., Tian, H., Wang, W., Zeng, N., Zhao, F., and Zhu, Q.: The North American Carbon Program Multi-Scale Synthesis and Terrestrial Model Intercomparison Project – Part 1: Overview and experimental design, Geosci. Model Dev., 6, 2121–2133, doi:10.5194/gmd-6-2121-2013, 2013.

Joiner, J.: GOME-2 version 26 (V26) 740 nm terrestrial chlorophyll fluorescence data, URL `https://acd-ext.gsfc.nasa.gov/People/Joiner/my_gifs/GOME_F/GOME-F.htm`, 2014.

Jones, R. H.: Bayesian information criterion for longitudinal and clustered data, Stat. Med., 30, 3050–3056, doi:10.1002/sim.4323, 2011.

King, A., Wullschleger, S., and Post, W.: Seasonal biosphere atmosphere CO_2 exchange and terrestrial ecosystem carbon storage: mechanism, extrapolation, and implications, in: Extended Abstracts of Fifth International Carbon Dioxide Conference. Cairns, Queensland, Australia, pp. 257–258, 1997.

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

Liu, J., Fung, I., Kalnay, E., and Kang, J.-S.: CO_2 transport uncertainties from the uncertainties in meteorological fields, Geophys. Res. Lett., 38, doi:10.1029/2011GL047213, l12808, 2011.

Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I., and Liu, J.: Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors, Atmos. Chem. Phys., 15, 2903–2914, doi:10.5194/acp-15-2903-2015, 2015.

Miyoshi, T.: The Gaussian approach to adaptive covariance inflation and its implementation with the Local Ensemble Transform Kalman Filter, Mon. Weather Rev., 139, 1519–1535, doi:10.1175/2010MWR3570.1, 2011.

Mueller, K. L., Gourdji, S. M., and Michalak, A. M.: Global monthly averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements, J. Geophys. Res. – Atmos., 113, doi:10.1029/2007JD009734, d21114, 2008.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cy., 24, doi:10.1029/2009GB003642, gB3023, 2010.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Ptron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Scie. USA, 104, 18 925–18 930, doi:10.1073/pnas.0708986104, 2007.

Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A.: Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, J. Geophys. Res. – Biogeo., 113, doi:10.1029/2007JG000603, g03034, 2008.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, URL `http://www.jstor.org/stable/2958889`, 1978.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014GL059684, 2014.

Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wofsy, S. C., and Xu, L.: A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS, Remote Sens. Environ., 112, 1633 – 1646, doi:10.1016/j.rse.2007.08.004, remote Sensing Data Assimilation Special Issue, 2008.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Change Biol., 9, 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

Tadić, J. M., Qiu, X., Miller, S., and Michalak, A. M.: Spatio-temporal approach to moving window block kriging of satellite data v1.0, Geosci. Model Dev., 10, 709–720, doi:10.5194/gmd-10-709-2017, 2017.

Tian, H., Melillo, J., Lu, C., Kicklighter, D., Liu, M., Ren, W., Xu, X., Chen, G., Zhang, C., Pan, S., Liu, J., and Running, S.: China's terrestrial carbon balance: Contributions from multiple global change factors, Global Biogeochem. Cy., 25, doi:10.1029/2010GB003838, gB1007, 2011.

Wu, C., Chen, J. M., and Huang, N.: Predicting gross primary production from the enhanced vegetation index and photosynthetically active radiation: Evaluation and calibration, Remote Sens. Environ., 115, 3424 – 3435, doi:10.1016/j.rse.2011.08.006, 2011.

Wylie, B. K., Johnson, D. A., Laca, E., Saliendra, N. Z., Gilmanov, T. G., Reed, B. C., Tieszen, L. L., and Worstell, B. B.: Calibration of remotely sensed, coarse resolution NDVI to CO_2 fluxes in a sagebrush-steppe ecosystem, Remote Sens. Environ., 85, 243 – 255, doi:10.1016/S0034-4257(03)00004-X, 2003.

Yadav, V., Mueller, K. L., and Michalak, A. M.: A backward elimination discrete optimization algorithm for model selection in spatio-temporal regression models, Environ. Modell. Softw., 42, 88 – 98, doi:10.1016/j.envsoft.2012.12.009, 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, doi:10.1002/2015GL063201, 2015GL063201, 2015.