Supplement of Atmos. Chem. Phys., 20, 323–331, 2020
https://doi.org/10.5194/acp-20-323-2020-supplement

Supplement of

The impact of improved satellite retrievals on estimates of biospheric carbon balance

Scot M. Miller and Anna M. Michalak

Correspondence to: Scot M. Miller (smill191@jhu.edu)

S1 The regression and model selection framework

We construct a series of statistical experiments to evaluate whether we can detect spatial and temporal variations in biospheric CO_2 fluxes within different global regions using current OCO-2 observations. This section of the SI describes these experiments in greater detail. This approach is identical to Miller et al. (2018), and that study provides additional detail.

These experiments are based upon a regression framework, as described in the main article. The regression has the following form:

$$\boldsymbol{z} \;=\; h(\mathbf{X})\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\epsilon} \tag{S1}$$

where $\boldsymbol{z}$ (dimensions $n \times 1$) are the OCO-2 observations, and $\mathbf{X}$ ($m \times p$) contains p different CO_2 flux tracers. These tracers include terrestrial biosphere model (TBM) estimates of CO_2 fluxes and remote sensing vegetation indices that are known to correlate with patterns in CO_2 fluxes (Sect. S2). There are different columns of $\mathbf{X}$ corresponding to each CO_2 flux tracer in each different month and each different global region; we run the regression on all months simultaneously. These tracers (both the TBMs and vegetation indices) are subsequently run through an atmospheric transport model $h()$, in this case the Parameterized Chemistry and Transport (PCTM) model (Kawa et al., 2004). The coefficients estimated as part of the regression ($\boldsymbol{\beta}$, $p \times 1$) scale these model outputs to best match the observations ($\boldsymbol{z}$). Furthermore, $\boldsymbol{b}$ ($n \times 1$) is the model spinup or CO_2 mixing ratios at the beginning of the experiments, and $\boldsymbol{\epsilon}$ ($n \times 1$) are the model–data residuals.

We pair this regression with model selection; model selection will determine which combination of model outputs (i.e., columns of $h(\mathbf{X})$) best describe variability in current OCO-2 observations. It will identify the set of model outputs with the greatest power to describe the data and ensures that the regression does not overfit the data (e.g., Zucchini, 2000). We specifically implement model selection based on the Bayesian Information Criterion (BIC), one of the most commonly-used forms of model selection (Schwarz, 1978; Mueller et al., 2010; Gourdji et al., 2012). We calculate a BIC score for many different combinations of model outputs, and each combination has a different set of columns ($h(\mathbf{X})$). The best combination has the lowest BIC score:

$$BIC \;=\; L + p\ln(n^*) \tag{S2}$$

where L is the log likelihood of a specific combination of model outputs, p is the number of model outputs in that combination, and n^* is the effective number of independent observations from OCO-2 during the study period. The log likelihood equation rewards combinations of model outputs that improve model-data fit, described in detail in Miller et al. (2018). By contrast, the second term in the equation ($p \ln n^*$) penalizes combinations with a greater number of model outputs, and it ensures that the selected combination is not an over-fit. This penalty and the log likelihood (L) not only depend upon the number of model outputs but also the effective number of independent observations (n^*). This number reflects the level of spatial and temporal correlation in the observational and model errors. If the spatial and temporal error correlations are large (i.e., bias-type errors), then the n^* will be small relative to the total number of OCO-2 observations. By contrast, if the errors are uncorrelated and completely independent, then n^* will equal the total number of OCO-2 observations. The companion paper Miller et al. (2018) describes in detail how we estimate this quantity.

S2 Additional detail on the tracers used in model selection

This section provides additional detail on the TBMs and vegetation indices that are used in the model selection experiments. These TBMs and vegetation indices are used as the input tracers in PCTM. We then generate forward atmospheric model simulations using PCTM and interpolate these model outputs to the locations and times of the OCO-2 observations to generate modeled XCO_2 total columns. These modeled columns become the columns of $h(\mathbf{X})$ in the model selection experiments (Eq. S1). Note that the multiple regression will scale the magnitude of the TBMs and vegetation indices 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 assesses the degree to which the spatial and temporal patterns in the TBMs and vegetation indices, after being transported through the atmosphere to the times and location of OCO-2 observations, can explain the spatial and temporal patterns in OCO-2 observations.

We include four TBMs from the recent Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP) (Huntzinger et al., 2013). These TBMs have very different space-time patterns and therefore represent a wide range of plausible flux patterns. The 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 products have a spatial resolution of 0.5° latitude by 0.5° longitude resolution, and we regrid these products to the PCTM model grid (1° latitude by 1.25° longitude). Furthermore, Fisher et al. (2016) downscaled the MsTMIP products to a 3-hourly temporal resolution; we use this version of the MsTMIP products in the present study.

The MsTMIP estimates are available through year 2010. Because these estimates are not available for the years of this study (2014-2015), we use a multi-year average as inputs in the PCTM model. Specifically, downscaled MsTMIP products are available from Fisher et al. (2016) for years 2004-2010, and we average the MsTMIP models over those years within each separate model grid box and for each separate 3-hourly time period to produce a multi-year average for each MsTMIP estimate. The resulting CO_2 flux estimates vary hour to hour and day to day but not year to year. Note that some recent inverse modeling studies using OCO-2 observations incorporate a prior flux estimate that has been generated for more recent years (e.g., Crowell et al., 2019). Unlike inverse modeling studies that often require a single prior flux estimate, we require numerous 3-hourly CO_2 flux tracers that represent a wide range of plausible patterns for the statistical model used in this study. The creation of a new, updated TBM inter-comparison is beyond the scope of this study. Furthermore, the objective of this study is to compare how the CO_2 flux constraint has improved as the retrievals have evolved from version 7 to versions 8 and 9. To facilitate this comparison, we have used the same set of flux models from MsTMIP as in the preceding companion paper (Miller et al., 2018).

In addition to these TBMs, we also utilize several vegetation indices as possible tracers of CO_2 fluxes within the regression ($\mathbf{X}$ in Eq. S1). These include the enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), and solar-induced fluorescence (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 SIF (e.g., Guanter et al., 2014; Yang et al., 2015; Shiga et al., 2018). These indices are therefore good candidate CO_2 flux tracers to use within the model selection experiments.

We specifically use EVI and NDVI estimates from the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua product MYD13C1 (Didan, 2015a) and the MODIS Terra product MOD13C1 (Didan, 2015b). These products are collectively available at 8-day intervals. Specifically, 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 also use level 2 SIF retrievals from the Global Ozone Monitoring Experiment-2 (GOME-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 to the PCTM model and incorporate the resulting model outputs as candidate variables in the $h(\mathbf{X})$ matrix.

S3 Differences in the model selection results for versions 8 and 9 of the observations

The model selection results using versions 8 and 9 of the observations are very similar but exhibit a few subtle differences (Fig. 3 and 4). Specifically, we select slightly fewer CO_2 flux tracers using version 9 than version 8 in Fig. 3. These small differences are due to the stochastic nature of the statistical model. The regression model used in this manuscript requires an estimate of error variances and estimates of the error correlation length and correlation time. These estimates are used to calculate n^* (Eq. S2). We estimate these parameters using a randomized sub-selection of the observations due to the very large size of the OCO-2 dataset, a procedure described in Miller et al. (2018). The results of the regression analysis therefore exhibit a subtle stochasticity depending upon which observations were randomly selected for the variance and covariance estimation. For example, when we re-run the analysis in Fig. 3, we sometimes select a flux model in one or two more months using version 9 relative to version 8 and sometimes obtain identical results using versions 8 and 9.

References

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, https://doi.org/10.1029/92JD00655, 1992.

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

Didan, K.: MYD13C1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 0.05 Deg CMG V006, Tech. rep., NASA EOSDIS Land Processes DAAC, https://doi.org/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, https://doi.org/10.5067/MODIS/MOD13C1.006, 2015b.

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, https://doi.org/10.5194/bg-13-4271-2016, 2016.

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

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, https://doi.org/10.1073/pnas.1320008111, 2014.

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, https://doi.org/10.5194/gmd-6-2121-2013, 2013.

Joiner, J.: GOME-2 version 26 (V26) 740 nm terrestrial chlorophyll fluorescence data, `https://acd-ext.gsfc.nasa.gov/People/Joiner/my_gifs/GOME_F/GOME-F.htm` (last access: 15 Nov. 2019), 2014.

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, https://doi.org/10.1029/2004JD004554, d18312, 2004.

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.

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

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, https://doi.org/10.1029/2009GB003642, gB3023, 2010.

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, https://doi.org/10.1029/2007JG000603, g03034, 2008.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, 1978.

Shiga, Y. P., Tadić, J. M., Qiu, X., Yadav, V., Andrews, A. E., Berry, J. A., and Michalak, A. M.: Atmospheric CO_2 observations reveal strong correlation between regional net biospheric carbon uptake and solar-induced chlorophyll fluorescence, Geophys. Res. Lett., 45, 1122–1132, https://doi.org/10.1002/2017GL076630, 2018.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1016/S0034-4257(03)00004-X, 2003.

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, https://doi.org/10.1002/2015GL063201, 2015GL063201, 2015.

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