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. 1). 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. 1). 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. 1). 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 section discusses how we characterize the variances (σ^2) and covariance structure ($\mathbf{V}$) of the model–data residuals ($\boldsymbol{\epsilon}$) (Eq. 3). An estimate of the variance is required to calculate the residual sum of squares (RSS, Eq. 6), and an estimate of the covariance structure is necessary to calculate the effective number of independent observations (n^*, Eq. 7).

We first describe our approach to 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 in the vicinity of each residual i, where i denotes an individual element of $\boldsymbol{\epsilon}$. The covariance structure likely differs in different locations and at different times (i.e., is non-stationary). 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 use a similar approach here.

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. 7). The denominator of Eq. 7 sums over each element i. For each element i in the summation, we use covariance parameters estimated for that element i. Note that Eq. 7 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. 7 only sums over 1000 elements. More precisely, Eq. 7 becomes

$$n^* = \frac{n}{1 + \frac{\sum_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} V_{i,j}}{1000}} \tag{S1}$$

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

The model selection equations (i.e., Eq. 6) also require an estimate for σ^2. We estimate a single value for σ^2 using all n residuals:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} \epsilon_i^2 \tag{S2}$$

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. 2.1 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 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):

$$\boldsymbol{\epsilon} \;=\; \mathbf{1}\beta_{0,r} + \mathbf{X_r}\boldsymbol{\beta_r} + \zeta \tag{S3}$$

where $\boldsymbol{\epsilon}$ are the model-data residuals from Eq. 1, $\beta_{0,r}$ 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. 2.1, 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 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 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.

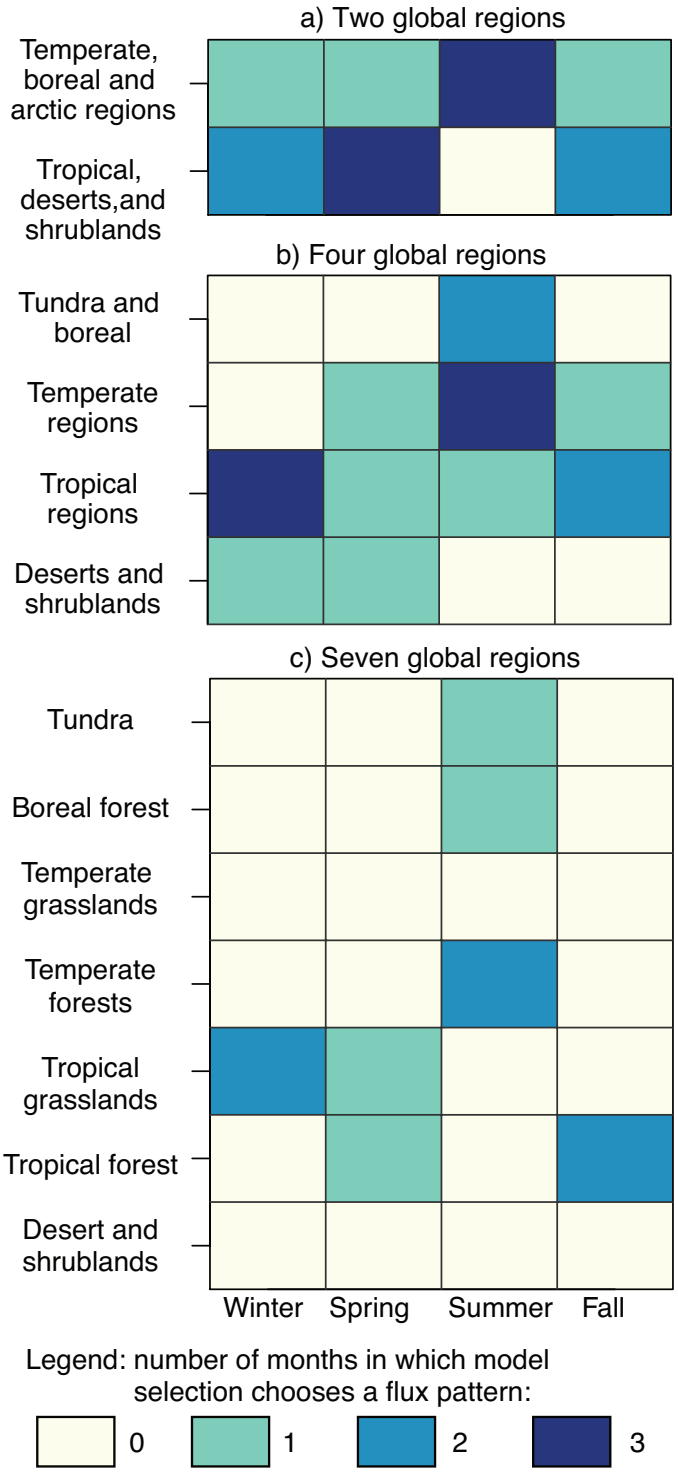

Figure S1: Results of the real data model selection experiment using all good quality OCO-2 retrievals (including glint mode retrievals). These model selection results are similar to those in the main manuscript that exclude glint mode data (Fig. 4).

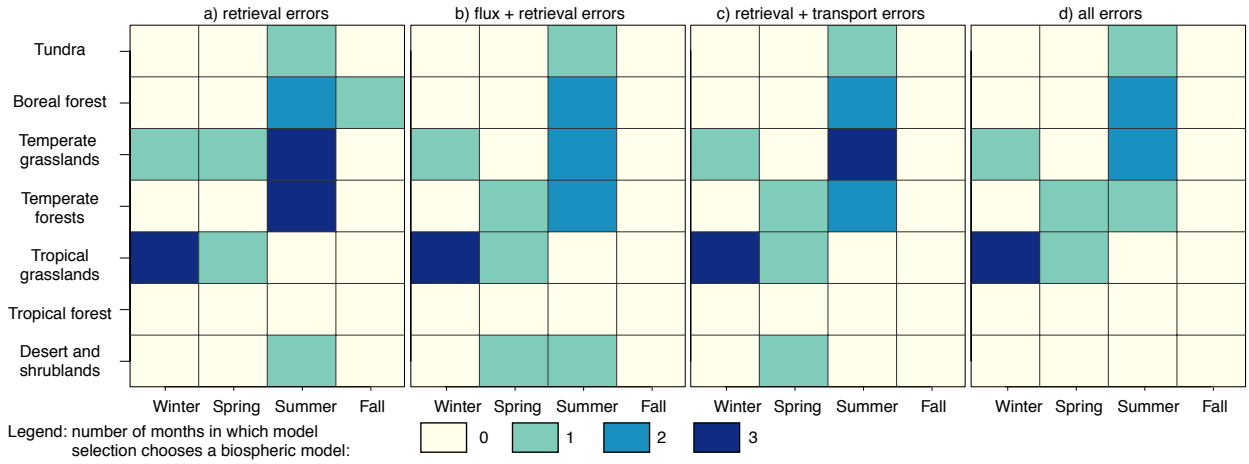

Figure S2: Results of the synthetic data case study using an alternative estimate for the retrieval errors (Sect. 2.1). These results are similar to those in Fig. 5, providing a consistency check on the synthetic data simulations.

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, doi:10.1029/92JD00655, 1992.

Didan, K.: MYD13C1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 0.05 Deg CMG V006, 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, 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.

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. Sci. 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` (last access: 30 Aug 2017), 2014.

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.

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. Sci. 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, Journal of Geophysical Research: Biogeosciences, 113, doi:10.1029/2007JG000603, g03034, 2008.

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.