The utility of continuous atmospheric measurements for identifying biospheric CO_2 flux variability

Deborah N. Huntzinger,[1] Sharon M. Gourdji,[1] Kimberly L. Mueller,[1] and Anna M. Michalak[1,2]

[1]   Motivated by the need to improve the modeling of land-atmosphere carbon exchange, this study examines the extent to which continuous atmospheric carbon dioxide (CO_2) observations can be used to evaluate flux variability at regional scales. The net ecosystem exchange estimates of four terrestrial biospheric models (TBMs) are used to represent plausible scenarios of surface flux distributions, which are compared in terms of their resulting atmospheric signals. The analysis focuses on North America using the nine towers of the continuous observation network that were operational in 2004. Four test cases are designed to isolate the influence on the atmospheric observations of (1) overall flux differences, (2) magnitude differences in flux across large regions, (3) differences in the flux patterns within ecoregions, and (4) flux variability in the near and far field of observation locations. The CO_2 signals generated from the different representations of surface flux distribution are compared using a Chi-square test of variance. Differences found to be significant are driven primarily by differences in flux magnitude over large scales, and the fine-scale (primarily temporal) variability of fluxes within the near field of observation locations. Differences in the spatial distribution of fluxes within individual ecoregions, on the other hand, do not translate into significant differences in the observed signals at the towers. Thus, given the types of variation in flux represented by the four TBMs, the atmospheric data may be most informative in the evaluation of aggregated fluxes over large spatial scales (e.g., ecoregions), as well as in the improvement of how the diurnal cycle of fluxes is represented in TBMs, particularly in areas close to tower locations.

1. Introduction

[2] Regional estimates of the imbalance in terrestrial carbon sources and sinks (net ecosystem exchange, NEE) generally have large associated uncertainties, due in part to the spatial complexity of the individual processes controlling carbon exchange at large scales. This is compounded by the fact that, at the global and regional scales, land-atmosphere carbon exchange cannot be measured directly [*Cramer et al.*, 1999]. Climate change predictions and carbon management decisions, however, depend on the ability to appropriately assess and model carbon uptake and release across various spatial scales. As a result, two main modeling approaches have been developed to estimate NEE at regional and continental scales: (1) terrestrial biospheric models (TBMs), which are based on current mechanistic understanding of how carbon is exchanged within ecosystems; and (2) atmospheric inverse models, which use measured atmospheric concentrations of CO_2, coupled with a transport model to infer surface flux distributions. In some cases, a combination of these approaches is used to optimize TBM parameters using atmospheric observations of CO_2.

[3] TBMs have become an integral tool for better understanding the mechanisms controlling carbon exchange across terrestrial ecosystems [*Waring and Running*, 2007]. Although TBMs can be used to link carbon sources and sinks to explicit ecosystem processes, they depend heavily on their simplifying assumptions, environmental driving data and initial conditions, as well as the way in which the processes controlling carbon exchange are formulated and scaled within the model. A particular model may do well at reproducing fluxes at a given location, however TBMs often differ considerably in their estimates of fluxes over larger regions [e.g., *Schimel et al.*, 1997; *Cramer et al.*, 1999; D. N. Huntzinger et al., North American Carbon Project (NACP) Regional Interim Synthesis: Terrestrial Biospheric Model Intercomparison, manuscript in preparation, 2011].

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA.
[2]Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan, USA.

In addition, current understanding of the processes controlling carbon exchange is not sufficient to rank models in terms of which is "best" at representing current fluxes or predicting carbon exchange under future climate conditions [*Heimann et al.*, 1998; *McGuire et al.*, 2001; *Melillo et al.*, 1995]. One way to assess model performance is to evaluate TBMs against atmospheric CO_2 observations, where an atmospheric transport model is used to transport surface fluxes to observation locations. The signal from the surface fluxes can be compared to the available observations, for example in terms of the depth and timing of the seasonal cycle, as well as interannual variations due to climatic factors, land-use change, and disturbances [*Heimann et al.*, 1998; *Nevison et al.*, 2008; *Dargaville et al.*, 2002; *Denning et al.*, 2003]. This type of evaluation approach depends, in part, on the ability of the atmospheric observations to detect differences between the surface flux distributions of various models.

[4] Conversely, flux estimates from atmospheric inverse models are more comprehensive, in the sense that all eco-system sources and sinks, fossil fuel emissions, and any other processes emitting or absorbing CO_2 are, in principle, captured in the atmospheric signal [*Global Change Project (GCP)*, 2010]. On the other hand, although inverse models predict fluxes that are quantitatively consistent with atmospheric measurements, atmospheric mixing, coupled with the sparseness of observations, leaves the problem ill-posed and frequently underconstrained [e.g., *Enting*, 2002]. Thus, multiple sets of surface flux estimates may be consistent with a single record of observed CO_2 concentrations [e.g., *Kaminski and Heimann*, 2001; *Enting*, 2002]. Generally, an additional constraint, such as explicit prior flux estimates from a TBM, must be included within the inversion to make the problem more tractable [e.g., *Kaminski et al.*, 1999; *Rödenbeck et al.*, 2003; *Gurney et al.*, 2004; *Baker et al.*, 2006]. In underconstrained regions, given the sparseness of the atmospheric network, many regional estimates from atmospheric inversions tend to revert to these explicit prior flux estimates, and, as a result, flux estimates can strongly reflect the characteristics of the specific TBM used within the inversion [e.g., *Gurney et al.*, 2003; *Butler et al.*, 2011]. Thus, the accuracy of inversion results depends on a number of factors, including the resolution (spatial and temporal) of the fluxes being estimated [*Kaminski et al.*, 2001; *Peylin et al.*, 2005], the accuracy of the transport model and prior estimates of flux [e.g., *Gurney et al.*, 2003; *Rödenbeck et al.*, 2003; *Baker et al.*, 2006], the density of the monitoring network, and the sensitivity of available atmospheric observations to the underlying flux distribution [e.g., *Gloor et al.*, 2001; *Gerbig et al.*, 2009; *GCP*, 2010]. To reduce aggregation errors, there has been an increasing focus in inversions on estimating fluxes at finer spatial and temporal resolutions in order to better resolve the potential responses of various vegetation types and the impact of human activities on regional carbon budgets [e.g., *Gerbig et al.*, 2003a; *Rödenbeck et al.*, 2003; *Michalak et al.*, 2004; *Peylin et al.*, 2005; *Peters et al.*, 2007; *Lauvaux et al.*, 2008; *Gourdji et al.*, 2010]. However, it remains uncertain, at finer spatial scales, whether a unique surface flux distribution can be derived from concentration measurements given the diffusive nature of atmospheric transport.

[5] The objective of the work presented here is to determine how much information atmospheric CO_2 observations can provide in either estimating surface flux distributions at regional scales (e.g., from inversions), or evaluating preexisting sets of surface flux estimates (e.g., from TBMs) across North America. This work is motivated by the need to improve both forward and inverse models. For TBMs, there is a need to validate flux estimates against observational data; thus, we examine the extent to which atmospheric CO_2 concentration data can be used to evaluate TBM model performance. In order to evaluate the relative merits of two or more TBMs, or to validate a single TBM using atmospheric data, however, the atmospheric data must be able to discriminate between models. Therefore, this manuscript evaluates whether atmospheric data can be used to do this, given the degree of variability between fluxes as estimated by different models and the errors associated with the ability to reproduce observations (i.e., model-data mismatch). Similarly, the recent focus in inverse modeling studies on finer-scale flux estimation raises questions about the ability of the atmospheric network to provide sufficient information to accurately infer fine-scale flux variations beyond those specified by any explicit prior flux estimates.

[6] To address the above objectives, the NEE estimates of four TBMs are used to represent plausible scenarios of surface flux distributions and magnitude, which are compared in terms of their resulting atmospheric signals. The analysis focuses on North America, and uses the nine towers of the 2004 continuous monitoring network in the analysis. Although the network has expanded since that time, the 2004 network is representative of the types of towers that are part of the larger, current observational network. Therefore, the 2004 network can be used to identify more general rules about what a tower is likely to "see" or detect, depending on its height, location, regional meteorology, and the ecoregions or land-cover types that surround it.

[7] Comparing the CO_2 signals originating from different surface flux distributions only indicates whether the concentration measurements, in general, can distinguish between different representations (e.g., model estimates) of flux. It does not provide information as to what is driving the differences (e.g., spatial or temporal variations in flux estimates, differences in flux magnitude), and to what extent areas far away from towers contribute to these differences. A biospheric modeler may be interested in not only knowing whether CO_2 measurements could be used to validate a given model, but over what regions such an analysis is appropriate and what type of information the CO_2 measurements can provide (e.g., spatial variability, temporal variability). Similarly, an inverse modeler may be interested in knowing how much information can be extracted from the CO_2 measurements, and how those measurements inform flux estimates over various spatial and temporal scales.

[8] Therefore, we use four (4) case studies to examine the spatial (and, to some extent, temporal) features of the distribution of surface carbon flux that can be detected by continuous tower observations of CO_2. Case 1 examines whether the atmospheric CO_2 measurements can detect overall differences between fluxes from different biospheric models, which is a necessary, but not sufficient, condition for evaluating fluxes with atmospheric CO_2 data. Then,

Table 1. The 2004 North American Continuous CO_2 Monitoring Network Tower Locations, Heights, and Estimated Model-Data Mismatch Variances

Tower Name	Location	Height (m)	Tower Type	σ_R^2 (ppm^2)
LEF	Park Falls, Wisconsin, USA	396	Tall	8.8
WKT	Moody, Texas, USA	457	Tall	5.9
SBL	Sable Island, Nova Scotia, Canada	25	Marine boundary	4.7
BRW	Barrow, Alaska, USA	10	Marine boundary	1.7
ARM	Norman, Oklahoma, USA	60	Short	11.7
HFO	Petersham, Massachusetts, USA	30	Short	38.6
AMT	Argyle, Maine, USA	107	Short	19.6
FRD	Fraserdale, Ontario, Canada	40	Short	7.9
CDL	Candle Lake, Saskatchewan, Canada	30	Short	4.0

Cases 2 through 4 involve manipulating the surface flux distribution estimated by the models to isolate the specific influence of ecosystem-scale variability, subecosystem-scale variability, and variability in the near versus far field of observation locations on CO_2 observations. Combined, the results from the different case studies are used to evaluate the extent to which atmospheric data can be used to evaluate and/or infer flux variability at various scales.

2. Methods

[9] To meet the objectives outlined earlier, synthetic CO_2 observations were generated for the highest sampling elevation of the nine towers that were collecting continuous CO_2 concentration measurements in North America during 2004 (Table 1). The highest measurement locations were used because they sample the most well mixed air and, thereby, better represent the influence of fluxes over the largest spatial footprint. These synthetic CO_2 signals were generated using different representations of surface carbon flux from four TBMs, and are influenced only by fluxes occurring within the domain of study. Therefore, there is no need to consider CO_2 concentrations of air coming into the North American domain, or the impact of fossil fuel emissions on the synthetic CO_2 signals. The signals were compared and evaluated using a combination of case studies (section 2.3) and statistical significance testing (section 2.4).

2.1. Sensitivity of CO_2 Observations to Surface Fluxes

[10] Atmospheric transport was modeled using the Stochastic Time-Inverted Lagrangian Transport Model (STILT) [*Lin et al.*, 2003] driven by analyzed winds from the Weather Research and Forecasting (WRF) model version 2.2 [*Skamarock et al.*, 2005]. STILT simulates the influence of upwind fluxes on observations by tracking the evolution of an ensemble of air parcels backward in time [*Lin et al.*, 2003]. The sensitivities of 3-hourly averaged atmospheric measurements at the towers to 3-hourly varying upwind surface fluxes at a 1° by 1° resolution over North America were derived following the methods outlined by *Lin et al.* [2003] and *Gourdji et al.* [2010]. Thus, the footprint describes how unit fluxes in a particular grid cell of the domain at a particular time affect the CO_2 concentration at the tower. The integrated sensitivities or footprints for each tower have units of ppm/(μmol/(m^{2}s)), and represent the influence of fluxes (in μmol m^{-2} s^{-1}) that occurred up to 10 days prior to the measurement on CO_2 concentrations (ppm) at the tower. An example of the combined footprint for all 9 towers is shown in Figure 1a.

2.2. Synthetic Observations

[11] The concentration footprints quantify the sensitivity of atmospheric observations to upwind fluxes, and are independent of the surface fluxes themselves. To simulate
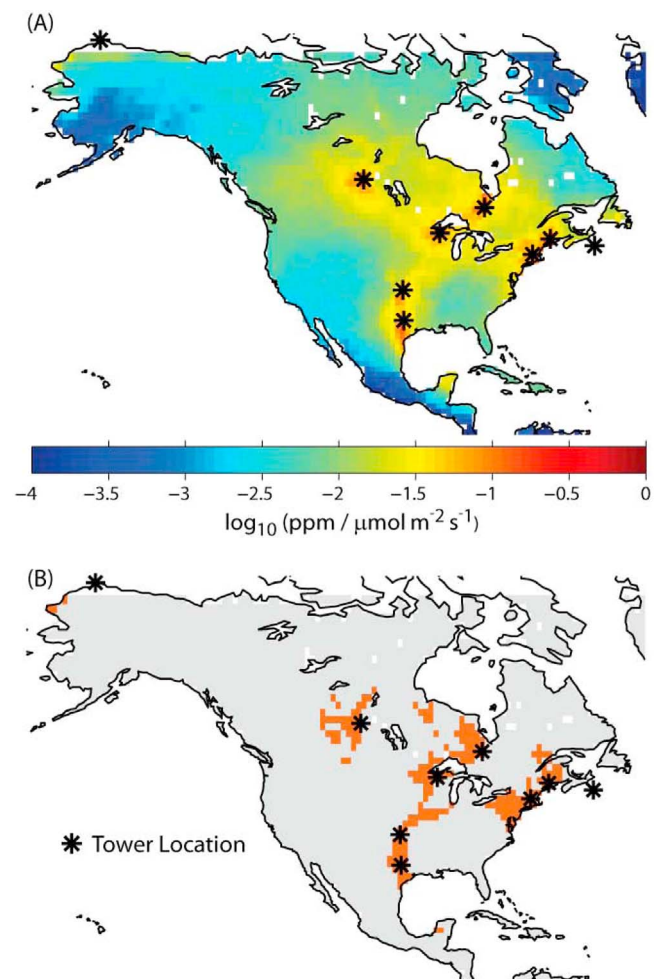

Figure 1. (a) Average sensitivity of July 2004 observations to surface fluxes. (b) Near field of measurement locations, defined as the areas to which the observations are most sensitive, with example shown for the week of 18 to 25 July 2004. High-sensitivity areas are defined as the cells that account for the upper 30% of the total sensitivity of the observations to the underlying fluxes.

synthetic CO$_2$ signals at the towers, the footprints were multiplied by a set of surface fluxes. Distinct representations of land-atmospheric flux distribution were defined using the 3-hourly, 1° × 1° surface flux distributions predicted by four TBMs: the Simple Biosphere Model (SiB3.0) [*Sellers et al.*, 1986; *Baker et al.*, 2008], Carnegie-Ames-Stanford-Approach as configured for the Global Fire Emissions Database v2 project (CASA GFEDv2) [*Potter et al.*, 1993; *Randerson et al.*, 1997; *van der Werf et al.*, 2006], Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) [*Krinner et al.*, 2005], and Vegetation Global Atmosphere and Soil (VEGAS2) [*Zeng*, 2003; *Zeng et al.*, 2005]. Given the aims of this study, the choice of TBM is somewhat subjective. However, these TBMs were chosen because they (1) provide distinct temporally and spatially variable maps of land-atmospheric surface flux; (2) have been widely applied across a variety of regions; and (3) have been used as prior estimates in inverse modeling studies [e.g., *Gurney et al.*, 2003; *Chevallier et al.*, 2006; *Peters et al.*, 2007; *Wang et al.*, 2007]. Monthly VEGAS2 and CASA GFEDv2 fluxes were temporally downscaled to 3-hourly resolution using the methods of *Olsen and Randerson* [2004] and net shortwave radiation and near-surface temperature from the NASA Global Land Data Assimilation System (GLDAS) [*Rodell et al.*, 2004]. The different flux representations from the four TBMs are shown in Figure 2, along with an example of the overall differences in their flux estimates, shown as the July average across-model standard deviation in 3-hourly fluxes.

[12] Three-hourly, synthetic observation signals were generated for each tower in Table 1. For the two tall and MBL towers, observations were generated throughout the day and night. Atmospheric transport models generally have difficulty in simulating the nocturnal planetary boundary layer (PBL) height [*Geels et al.*, 2007], which can result in biased flux estimates when using nighttime data. Therefore, following *Carouge et al.* [2010] and *Gourdji et al.* [2010], only afternoon (1800–2400 UTC) footprints were used to generate concentrations at the shorter towers (≤100 m) (Table 1).

[13] Although errors in the transport model may impact the resultant concentrations, the same atmospheric transport model (i.e., same footprints) is used to create the four synthetic concentration signals, and differences in the signals are, therefore, due solely to the underlying fluxes. While a different transport model could yield slightly different conclusions, WRF/STILT was selected here because it has been applied in several studies aimed at estimating CO$_2$ sources and sinks within North America [e.g., *Gerbig et al.*, 2003b; *Gourdji et al.*, 2010; *Lin et al.*, 2004].

2.3. Case Studies

[14] Four test cases were designed to isolate the influence on the atmospheric measurements of: (1) overall flux differences (Case 1); (2) magnitude differences in flux across large regions (Case 2); (3) differences in flux pattern within ecoregions (Case 3); and (4) flux variability in the near versus far field of measurement locations (Case 4). Cases 2 through 4 require manipulation of the surface flux distributions from the four TBMs in order to isolate specific influences, as described below and illustrated in Figure 3.

2.3.1. Case 1: Differences in Both the Distribution and Magnitude of Fluxes

[15] Case 1 examines the overall combined influence of surface flux magnitude and spatial distribution on CO$_2$ concentrations by using the unique flux distribution from each TBM to generate synthetic observations at the 9 tower locations (Figure 3a). As such, Case 1 examines whether the concentration measurements resulting from flux distributions as estimated by different TBMs are different from one another. Cases 2 through 4 are then used to determine the relative importance of various components of across-model flux variability on the observed CO$_2$ concentration variability in the synthetic observation signals.

2.3.2. Case 2: Subecoregion-Scale Variability Removed

[16] To isolate the influence of differences in regional flux magnitude on the generated CO$_2$ signals, the flux distributions of the TBMs are normalized to remove subecosystem-scale variability, i.e., the spatial variability within each ecoregion (Figure 3b). North America is divided into spatially contiguous ecoregions based on the work of *Olson et al.* [2001] and the model-specific weekly mean flux is calculated for each ecoregion and TBM. These mean fluxes are then applied to every 3-hourly period and 1° by 1° cell within that ecoregion over the weekly period, thereby removing subecoregion-scale spatial and subweekly temporal variability, while preserving differences in regional flux magnitude.

2.3.3. Case 3: Normalized Net Ecoregion-Scale Flux

[17] To examine how the distribution of fluxes within ecoregions influences CO$_2$ concentrations, the flux distribution of each TBM is normalized to have the same net area-weighted weekly flux by ecoregion; however, the distribution of fluxes within each of these ecoregions remains unique to each model (Figure 3c); that is, the 3-hourly temporal and 1° by 1° spatial variations in surface flux remain defined by the TBMs.

2.3.4. Case 4: Isolating Influence of Near- Versus Far-Field Fluxes

[18] The near field and far field are terms used in the literature to refer to those areas within close proximity to atmospheric observations and those areas that are farther away, respectively. For example, *Gerbig et al.* [2009] defines the near field as the area within about 50 km from the measurement location. Their work indicates that the spatial variations of surface fluxes in the near field of measurement locations contribute significantly to the observed variability in CO$_2$ concentrations [*Gerbig et al.*, 2003b, 2009]. While observations may be more sensitive to the spatial and temporal variability of fluxes within the near field of tower locations, the ability of the observations to detect differences in the spatial distribution of fluxes in the far field is less well known. Thus, Case 4 was designed to assess the impact of the subecosystem-scale variability beyond the near field of the towers.

[19] Here, the near field is defined as those 1° by 1° grid cells contributing to the greatest or upper 30% of the sensitivities of the atmospheric observations (Figure 1b), resulting in a near field that is much larger than that discussed by *Gerbig et al.* [2009]. In contrast to the high-resolution grid (20 km) near the towers used by *Gerbig et al.* [2006, 2009], here we examine the influence of 1° by 1° regional flux variations on CO$_2$ concentrations. Thus, the
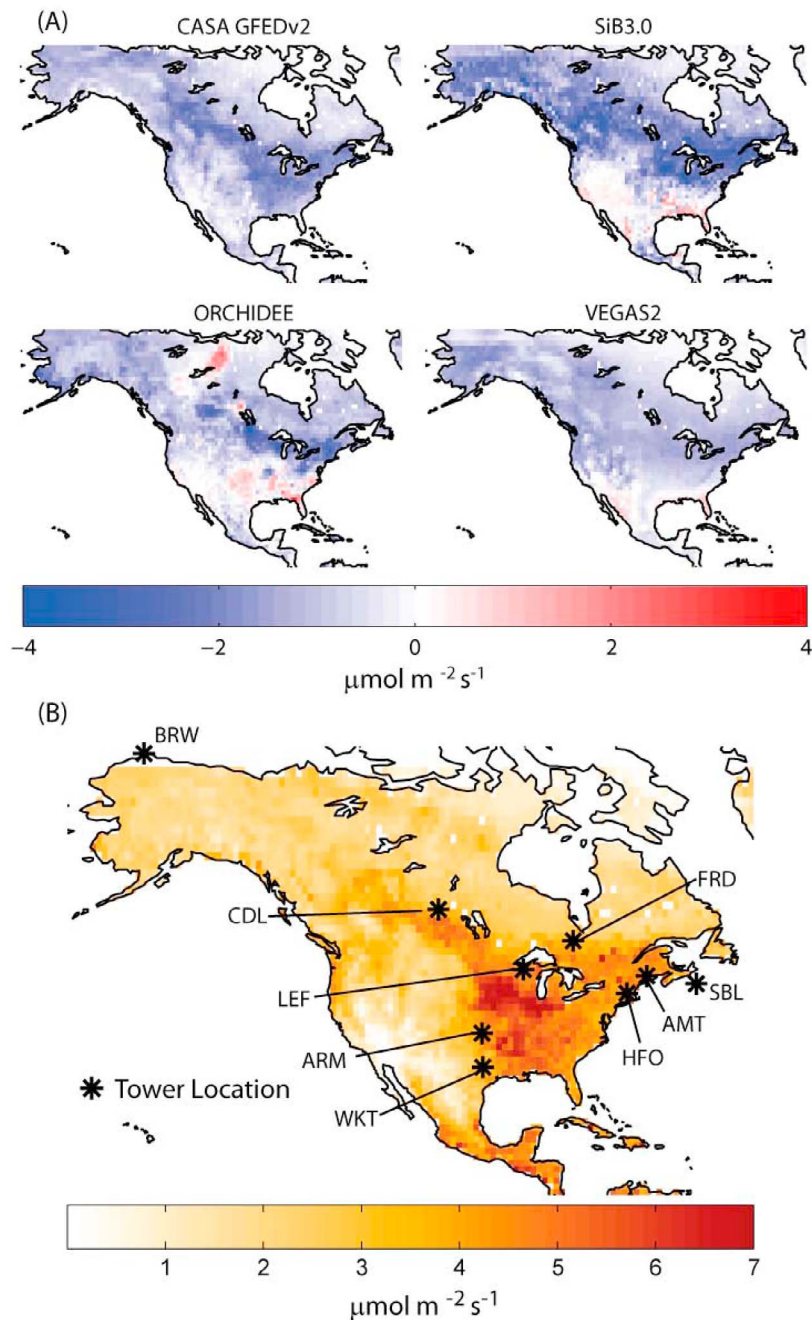
Figure 2. (a) Mean 2004 summer (June, July, and August) net ecosystem exchange (NEE) from the four biospheric models used to generate different "representations" of the surface flux distribution. (b) Average across-model standard deviation in NEE estimates for all 3-hourly periods in July 2004. Tower labels are defined in Table 1.

definition of the near field must be based on this grid. Furthermore, depending on the wind regime and the height of the boundary layer, a given measurement location is sensitive to different regions (i.e., fluxes) over time. As a result, the spatial extent and shape of the near field is expected to change with atmospheric transport. Therefore, here, the cells defining the near field of observation locations are allowed to vary by week, and we define the near field in terms of those areas to which the observations are most sensitive, rather than specifying a fixed area or distance from the tower.

[20] For each tower and each week, the land cells are sorted in terms of how sensitive the tower measurements are to surface fluxes from that cell. Cells are ordered from those having the greatest influence to those having the least. The influences (i.e., sensitivities) are cumulatively summed and divided by the overall or total sensitivity for that tower to the entire domain. Cells that contributed to the top 30% of the
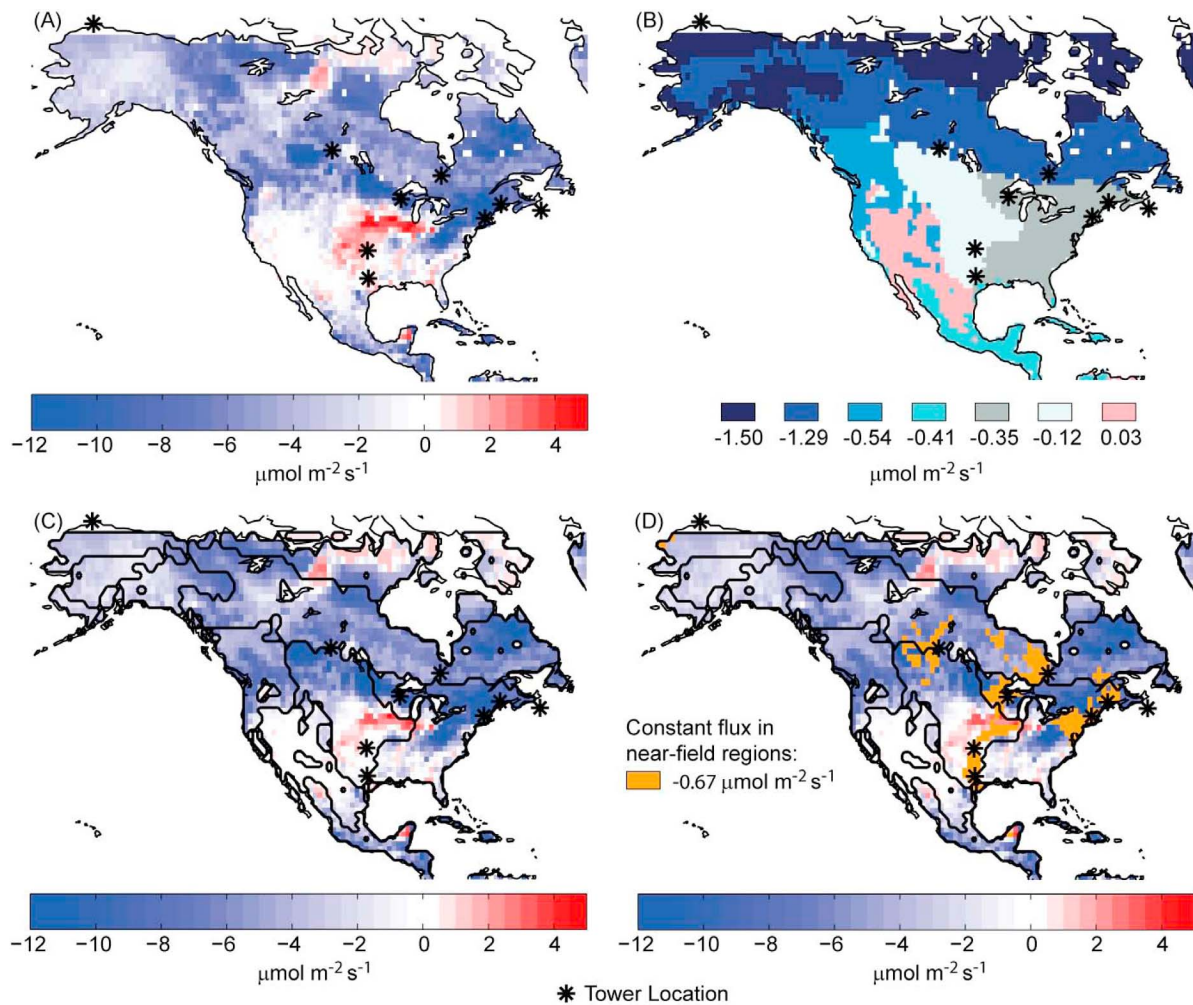
Figure 3. Examples of the different test cases using fluxes from ORCHIDEE for the week of 18 to 25 July 2004. (a) Case 1: NEE as represented by the biospheric model. (b) Case 2: different flux magnitude by ecoregion among the models, but with subecoregion-scale variability removed. (c) Case 3: same net flux by ecoregion across models, but with the distribution of fluxes within each ecoregion remaining unique to each model. (d) Case 4: Near field assigned a constant flux equal to the across-model mean for that area, while preserving, for each model, the spatial distribution of fluxes in the far field as defined in Case 3.

sensitivity are defined as the near field for that tower and that week. Figure 1b shows an example of the combined near field for all of the 9 towers in the 2004 network. The far field is defined as all other land cells outside the near field. Based on the above criteria for defining the near field, the near field constitutes, on average, 14% of the land cells in North America.

[21] To isolate the impact of the far field, an across-model, weekly mean is applied to every 3-hourly period and cell within the near field of the tower or observation locations. The subecosystem and temporal variability of the far field is then defined as described in Case 3 (Figure 3d).

2.4. Significance Testing

[22] One of the goals of this analysis is to examine the potential of using real atmospheric CO_2 measurements to help validate or compare different TBMs. However, even if

surface fluxes were perfectly known, a mismatch between the modeled (i.e., surface flux convolved with an atmospheric transport model) and observed CO_2 observations is expected, termed "model-data mismatch" in the inverse modeling literature [e.g., *Bousquet et al.*, 1999; *Gurney et al.*, 2002; *Peylin et al.*, 2002; *Michalak et al.*, 2005]. This mismatch is primarily due to transport model errors, but also includes aggregation, representation, and measurement errors [*Kaminski et al.*, 2001; *Engelen et al.*, 2002], as well as possible errors introduced by choices in the inversion setup (e.g., correlation structure of fluxes).

[23] In order to assess whether the differences between the synthetic CO_2 concentrations modeled in this analysis using the different TBMs are significant, the synthetic signals are compared within the context of expected or estimated model-data mismatch error. Thus, the analysis takes into account the sources of error described in the previous par-

agraph, and the impact these uncertainties can have on the resultant CO_2 signals seen at the towers. In essence, comparing signals within the context of the expected model-data mismatch errors makes it possible to assess how different the signals have to be before this difference can confidently be attributed to differences in flux rather than errors inherent in the system. Although the current study factors out model data mismatch errors by using the same transport model to generate synthetic measurements for the four TBMs, the resulting signals are still compared in light of the types of uncertainties that would be present in comparisons with real atmospheric data. Model-data mismatch can be estimated in a number of ways, such as obtaining values from the literature coupled with analysis of the atmospheric measurements [e.g., *Schuh et al.*, 2010, *Butler et al.*, 2011], assessing the temporal variability in the data around a smoothed curve and assuming that this variability is representative of model-data mismatch errors [e.g., *Bousquet et al.*, 1999; *Gurney et al.*, 2002], or by using a statistical optimization method with the atmospheric measurements [e.g., *Michalak et al.*, 2005; *Gourdji et al.*, 2010]. A more detailed summary of how model-data mismatch variance is quantified in inversion studies is provided by *Michalak et al.* [2005].

[24] In order to obtain realistic estimates of model-data mismatch error, it is necessary to use real atmospheric measurements. Because this study compares synthetic concentration signals generated using a single transport model, using real atmospheric concentration measurements is the only way to estimate transport model error and any aggregation error below the 3-hourly and 1° by 1° temporal and spatial resolutions. For the current study, we use 3-hourly real atmospheric observations taken during the 2004 growing season at the towers included in this study (Table 1), and optimize the model-data mismatch (σ_R^2) using the Restricted Maximum Likelihood (RML) approach as implemented by *Gourdji et al.* [2010] at a 3-hourly flux resolution. This provides a conservative estimate of model-data mismatch variances for biospheric fluxes, because model-data mismatch evaluated using real data also includes the influence of uncertainty in fossil fuel emissions. Although the model-data mismatch variances are specific to WRF/STLT as estimated using the RML approach, we expect these to be similar to variances obtained for other contemporary atmospheric transport models.

[25] In general, model-data mismatch tends to be relatively low at tall and marine boundary layer towers (Table 1), where the air is relatively well mixed, whereas model-data mismatch can be much higher at shorter towers where local influences tend to have a greater influence on measured concentrations. This is particularly true of shorter towers close to areas with large variability in the flux distribution.

[26] Differences among pairs of synthetic observation signals from different TBMs are quantified using their mean squared difference (MSD). The MSD incorporates both the variance and the bias or offset between the two signals, and thus quantifies how different, on average, the model signals are from one another. As such, it is directly comparable to model-data mismatch variance, σ_R^2, or how different, on average, we would expect modeled CO_2 concentrations to be from the true concentrations. If the MSD between two synthetic concentration signals is greater than the estimated model-data mismatch variance (σ_R^2) at a given tower, then differences in the underlying flux distribution are more likely to be detectable by the real atmospheric observations at those towers. And, being able to detect these differences is important for distinguishing between competing biospheric models.

[27] The statistical significance of the difference between pairs of synthetic observation signals is quantified using an upper-tailed, Chi-square test of variance, where the Chi-square test statistic (χ^2) is defined as

$$\chi^2(t) = \frac{\nu \, \text{MSD}}{\sigma_R^2}, \qquad (1)$$

where ν is the number of degrees of freedom (n-1), n is the number of 3-hourly synthetic concentration observations over the examined period t at a particular tower, MSD (ppm^2) is the mean squared difference in 3-hourly concentrations between a pair of the synthetic observation time series over that same period, and σ_R^2 (ppm^2) is the model-data mismatch variance for a given tower.

[28] The analysis is performed monthly, on all possible pairs of synthetic observations in each test case. Statistical significance is determined within a hypothesis testing framework, where the null hypothesis is that the MSD between signals at a given tower is equal to the estimated model data mismatch variance (σ_R^2) for that tower. The alternative hypothesis is that the MSD between the signals is greater than σ_R^2. For each pair of signals, a p value for the test statistic (χ^2) is calculated using a Chi-square distribution with ν degrees of freedom. The p value represents the significance of the difference between two synthetic concentration signals, and, therefore, the lower the p value, the more significant the result. Thus, for each time period, the test establishes the likelihood that the difference between any two synthetic CO_2 signals would be detectable.

3. Results and Discussion

[29] The weekly averaged synthetic CO_2 signals generated using the different TBMs for the examined test cases are shown for four of the nine towers in the 2004 North American continuous atmospheric monitoring network in Figure 4. These four towers were chosen to provide an example of a tall tower (LEF); a marine boundary layer tower (SBL); a tower in an agricultural region (ARM); and one located in a highly productive and spatially variable forested region close to heavily populated areas with large fossil fuel sources (HFO) (see Table 1 for details). Synthetic observations at the temporal resolution used in the case studies (3-hourly) are presented for July 2004 in the auxiliary material Figure S1.[1] The results from each case study are discussed below, along with the implications of these results for the evaluation and comparison of TBM estimates using atmospheric data, as well as the potential implications of the results for the inference of fluxes through atmospheric inversions.

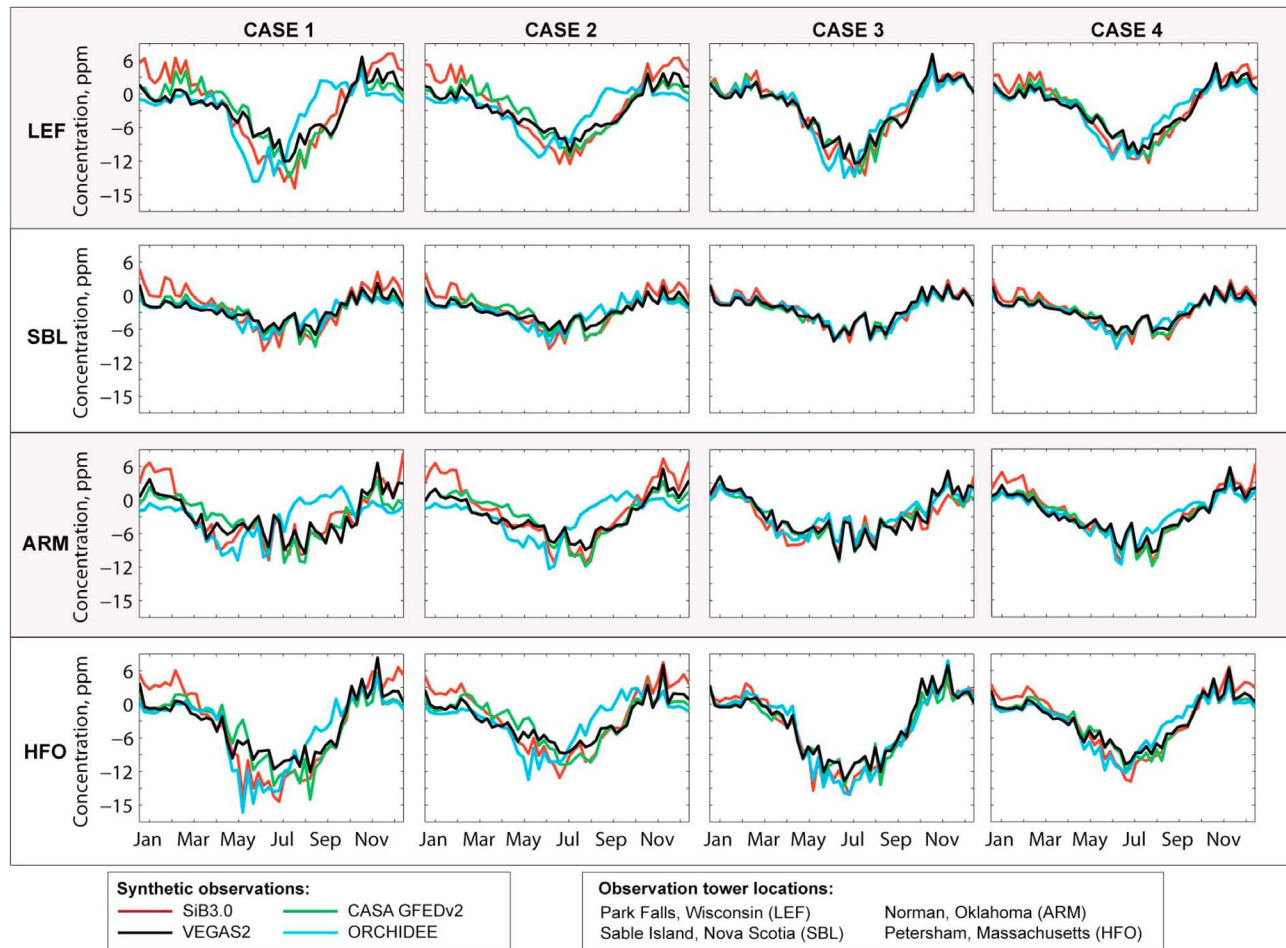[1]Auxiliary materials are available in the HTML. doi:10.1029/2010JD015048.

Figure 4. Weekly average synthetic observation signals for each test case, for four of the nine towers in the 2004 monitoring network (see Table 1). Towers were chosen to represent a tall tower (LEF), a marine boundary tower (SBL), a tower in an agricultural region (ARM), and a tower in a forested region (HFO).

3.1. Case 1: Differences in Both the Distribution and Magnitude of Fluxes

[30] Case study 1 is designed to assess the ability of the atmospheric network to detect differences in the CO$_2$ signal resulting from different representations of surface flux (section 2.3, Figure 3a). If the synthetic signals generated from these different flux distributions are not statistically significantly different from each other, then atmospheric CO$_2$ measurements may not be useful in evaluating or comparing flux estimates from biospheric models. It would also suggest that inversions cannot be used to infer the types of differences in fluxes represented by the TBMs examined in this study.

[31] The unique flux distributions from the four TBMs yield different weekly averaged synthetic observation signals, both in terms of the depth and timing of their seasonal cycles (Figure 4). Using 3-hourly concentrations (e.g., auxiliary material Figure S1), the six possible signal pairs are compared to evaluate the overall differences in atmospheric signals generated by the examined TBMs. On a weekly basis, the MSD between the 3-hourly synthetic CO$_2$ time series among these pairs is quite variable (Figure 5).

The significance of these differences is examined on a monthly time interval using the model-data mismatch variance estimated from real-concentration data as described in section 2.4. At most towers, the four TBMs generate statistically different 3-hourly synthetic CO$_2$ time series during most months of the year (Figure 6). The differences among the synthetic CO$_2$ signals are less significant in the winter months where there is less temporal variability in the 3-hourly concentrations (and fluxes), and smaller overall differences in the spatial distribution of fluxes among the models. Differences between synthetic signals are not as significant at BRW, where the land cells surrounding the tower generally have a smaller influence on atmospheric CO$_2$ concentrations measured at the tower (e.g., Figure 1). Differences are also less detectable at towers with a higher model-data mismatch variance. Towers such as HFO and AMT are located in areas with stronger variations in fluxes (e.g., highly productive forested regions or towers close to large fossil fuel sources) and tend to have a higher model-data mismatch variance due to greater representation and/or aggregations errors [*Kaminski et al.*, 2001; *Gerbig et al.*, 2009; *Gourdji et al.*, 2010].
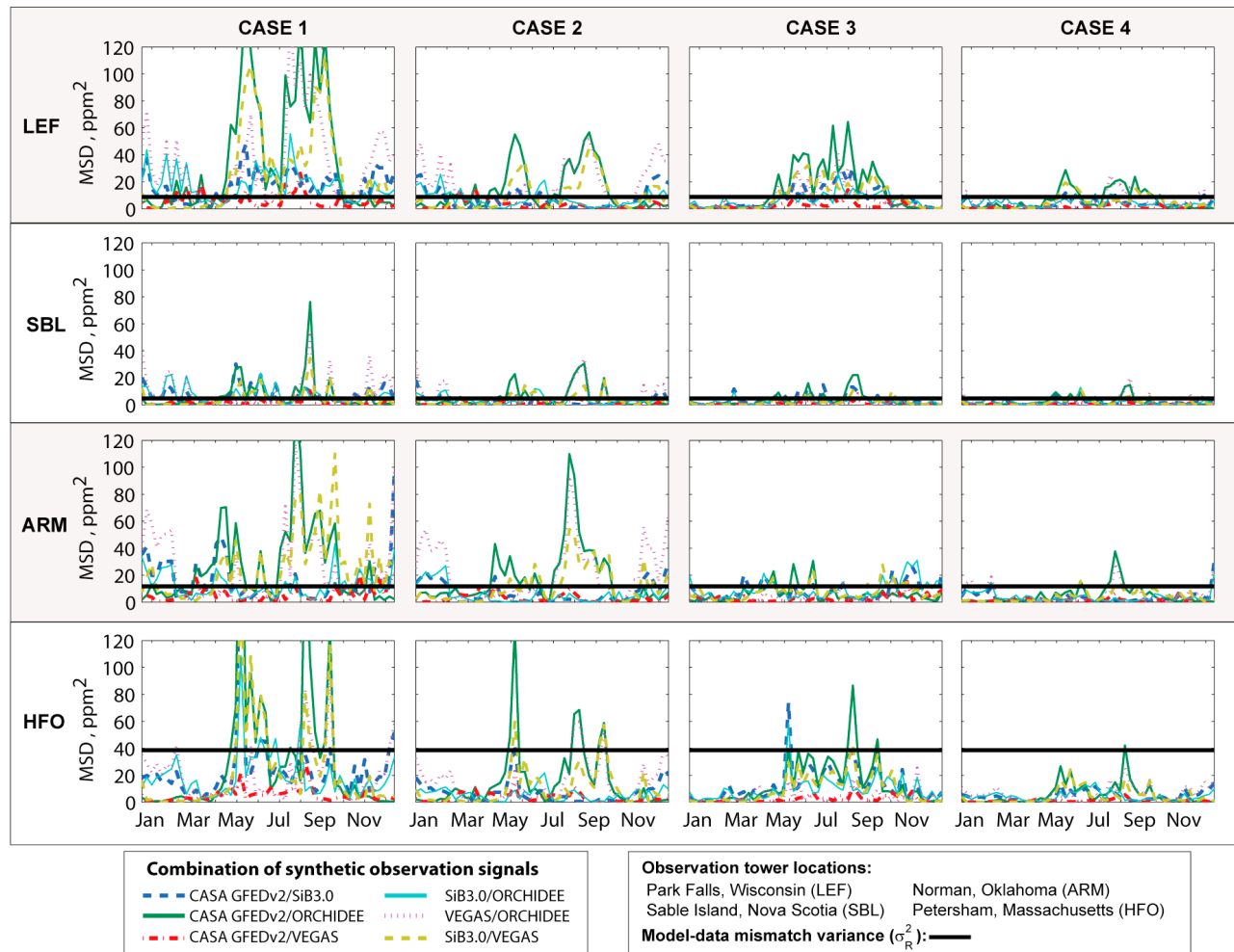
Figure 5. Weekly mean squared difference (MSD, ppm^2) between the six combinations of 3-hourly synthetic observation signals. The estimated model-data mismatch variance (σ_R^2, ppm^2) for four of the nine towers in the 2004 monitoring network (see Table 1) is shown in black. Towers were chosen to represent a tall tower (LEF), a marine boundary tower (SBL), a tower in an agricultural region (ARM), and a tower in a forested region (HFO). When the MSD between the signals is above the black line, the tower would be more likely to detect the difference between flux signals, taking into account measurement, transport, and representation errors.

[32] The magnitude of the estimated model-data mismatch variance has a direct impact on the significant testing results. This can be visualized by imagining moving the black line up or down in Figure 5 while keeping the MSD among the six pairs of synthetic signals constant. The closer the black line moves to zero, the more often the signal differences exceed the model-data mismatch variance, and therefore, the more often these differences would be detectable by the measurements. For a given grid or spatial resolution of fluxes, it is expected that as the ability to model atmospheric transport improves (i.e., σ_R^2 decreases), differences in atmospheric concentrations, and therefore the underlying fluxes, will become more detectable.

[33] In general, the results from Case 1 are encouraging, in that the atmospheric measurements can detect overall differences in CO$_2$ concentrations resulting from competing flux distributions (e.g., different TBMs). The differences in flux among the TBMs examined in this analysis are assumed to be comparable to the discrepancy between some "true" flux distribution and flux as represented by a given TBM. The degree to which the atmospheric data can be used to evaluate models (e.g., Is one model better than another? Are the fluxes predicted by a given model compatible with the atmospheric observations?) depends on whether the atmospheric data can detect differences among competing flux distributions. The ability to evaluate models also depends on the scales for which the atmospheric data are most informative about the underlying flux distribution (e.g., close to the tower, large regional flux differences). Below, the remaining cases examine how the temporal and spatial variability in surface flux translates into the variability observed in atmospheric CO$_2$ concentration seen in Case 1.

3.2. Case 2: Subecoregion-Scale Variability Removed

[34] Land-atmosphere carbon fluxes exhibit variability at different scales. Case 2 examines the ability of atmospheric
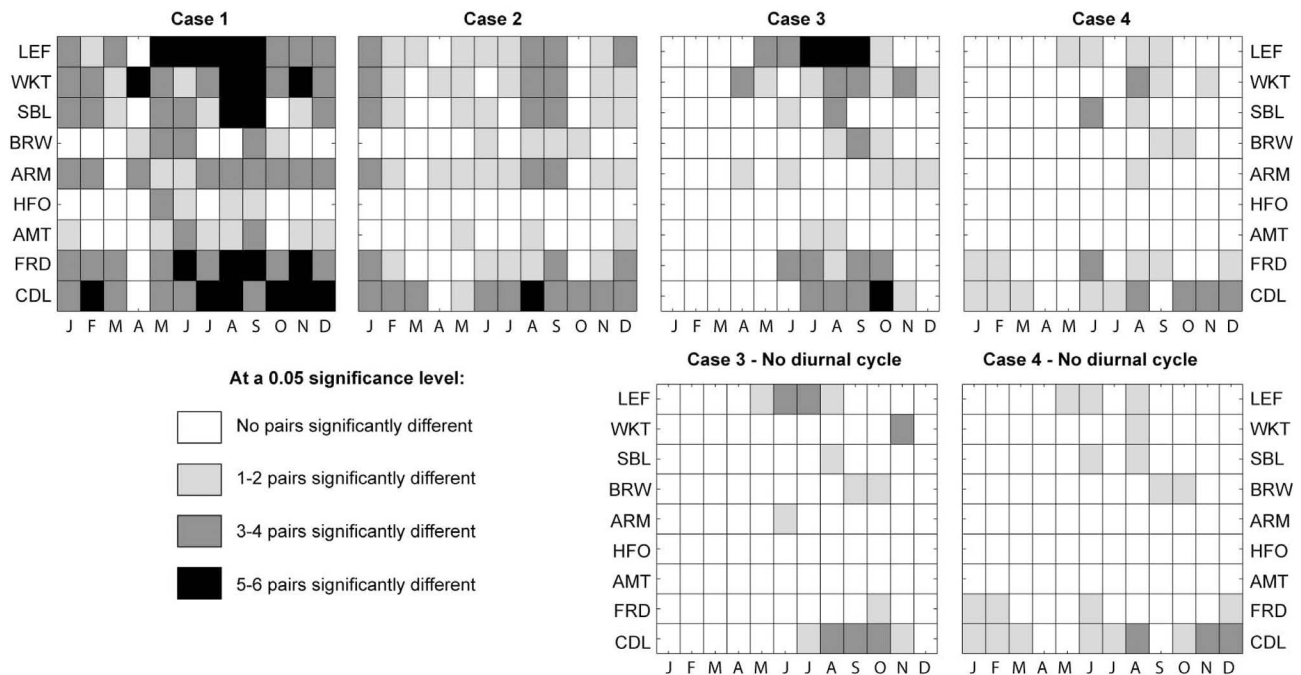
Figure 6. Statistical significance of the difference between the 3-hourly synthetic observation signals for different months, given the estimated model-data mismatch for each tower (Table 1). Shading indicates the number of signal pairs where the mean squared difference (MSD) between the synthetic observation signals is significantly above the model-data mismatch variance at a significance level or α value of 0.05. Towers are as defined in Table 1.

concentration measurements to detect weekly differences in large-scale (net) carbon flux (section 2.3, Figure 3b). A weekly time period was chosen for averaging fluxes to preserve the large-scale seasonal cycle of fluxes, while isolating magnitude differences in flux across large regions.

[35] Overall, the towers are able to detect differences in flux magnitude for large regions (Figure 5). However, the MSD between the 3-hourly concentration signals in Case 2 is smaller than observed in Case 1 (Figure 5). As a result, fewer pairs of synthetic CO_2 signals are significantly different from one another, particularly in the late winter and early spring months (February through May). However, in early winter (December and January) and later summer (August and September), the differences in regional-scale net flux between the TBMs have a greater impact on tower observations relative to other months. These seasonal variations in differences are likely due to the larger tower footprints in the winter months (i.e., the towers sampling air from a larger region because the PBL may be below the sampling height at taller towers and/or because wind patterns change more frequently in the winter thereby allowing a given tower to sample a larger area), as well as larger net differences in flux during August and September among the models.

[36] The results from Case 2 indicate that differences in modeled flux magnitude over large ecoregions are generally detectable by atmospheric measurements, and these observations can therefore be used, for at least some regions and seasons, to discriminate among large-scale fluxes as predicted by different TBMs. This is also encouraging for atmospheric inversions that estimate fluxes over large

regions, because the atmospheric data appear to provide sufficient information to help constrain fluxes at these scales. In fact, many inversions prescribe flux patterns within large regions (similar to the ecoregions used in this study), and adjust an initial or a priori estimate of the overall magnitudes of fluxes at the regional scale [e.g., *Peylin et al.*, 2001; *Law et al.*, 2002; *Peters et al.*, 2007].

[37] Using inversions to scale fluxes for large regions assumes that the atmospheric data are providing information about the large-scale regional variations in flux. Due to spatial aggregation errors, however, there is a debate within the inversion community about the number of estimation regions to use, or whether they should be used at all (e.g., estimate fluxes at grid scale rather than by region) [e.g., *Kaminski and Heimann*, 2001; *Kaminski et al.*, 2001; *Peylin et al.*, 2001]. Although the atmospheric data appear to provide information about large-scale fluxes that can be used to scale regions, any errors in the prescribed sub-ecosystem-scale variability of the a priori fluxes within the regions can lead to significant errors in the inversion result, particularly in regions close to the tower. Cases 3 and 4 evaluate the influence of surface flux variability and the scales (near versus far field) at which flux differences are detectable by the atmospheric data.

3.3. Case 3: Normalized Net Ecoregion-Scale Flux

[38] Case 3 is designed to evaluate whether the atmospheric network can detect difference in CO_2 observation signals caused by how fluxes are distributed within ecoregions (section 2.3, Figure 3c). If substantially different source/sink configurations do not yield substantially different synthetic

CO_2 time series, then the atmospheric data cannot be used to evaluate the spatial distribution (e.g., $1° \times 1°$ in this study) of fluxes from TBMs. The implications for inversions are slightly more complicated, particularly when the flux distribution of the prior estimate is prescribed or fixed by the inversion, as with inversions that estimate fluxes for regions (e.g., ecoregions, continents) larger than the resolution of the atmospheric transport model. For example, if the flux patterns within ecoregions do not significantly impact atmospheric observations, then, although an inversion that prescribed patterns based on a particular prior set of fluxes is able to reproduce observations, this does not imply that the prior had the correct patterns. This would imply that the results of such an inversion could not be interpreted at subecoregion scales.

[39] Results from Case 3 suggest that the atmospheric measurements can detect differences in the 3-hourly synthetic signals due to subecoregion-scale differences in the flux distribution (Figures 5 and 6) for some seasons and regions. Differences among the signals are seen primarily during the growing season (May through September), when both the magnitude and spatial heterogeneity of fluxes is greatest, and particularly in forested and northern regions of North America. The towers with larger summertime footprints (e.g., CDL, FRD, LEF) tend to have greater overall observation sensitivities to surface fluxes in regions farther from the tower location (Figure 1). A combination of larger tower footprint, along with higher across-model differences in surface flux distributions (Figure 1b) contribute to greater differences in the synthetic signals at these towers (Figure 6).

[40] While the results from Case 3 indicate that the atmospheric measurements are able to detect differences in the fine-scale (spatial and temporal) variability of flux, what is not clear is the relative importance of the spatial distribution of fluxes versus differences in the diurnal cycle of those fluxes (e.g., timing, strength). In order to further investigate the impact of differences in fine-scale flux distributions, Case 3 was repeated with the diurnal cycle of fluxes removed. Thus, the distribution of fluxes within each ecoregion remained unique to each model, but the temporal variability contained in the diurnal cycle was removed. When diurnal variability is removed, the significance of the difference between signals disappears for many model pairs and towers (Figure 6). Thus, the differences in signals observed in Case 3 are driven primarily by differences in the diurnal cycle of fluxes between the TBMs (original Case 3) rather than by differences in how the fluxes are spatially distributed among the models (modified Case 3).

[41] The impact of the diurnal cycle prescribed by the TBMs on the differences seen in the synthetic CO_2 signals at the towers has several potential implications for atmospheric inversions. Most importantly, if the magnitude and/or timing of the diurnal cycle in any prescribed fluxes are incorrect in atmospheric inversions, significant errors could be introduced into the inversion results. In order to avoid large temporal aggregation errors (and biases), inversions that use continuous observations have to account for the diurnal cycle in some way [*Law et al.*, 2004]. For example, the inversion could be allowed to adjust the diurnal variability in fluxes in order to properly account for the observed high-frequency variability in concentrations [e.g., *Gourdji et al.*,

2010], or the diurnal variability in the prior could simply be assumed to be correct. The latter approach is the most common among inversions, and is often implemented by either subtracting from the atmospheric observations the signal generated from a forward simulation run with the prior that includes the diurnal cycle [e.g., *Peters et al.*, 2007], or relying by on the prior to define the high-frequency time variations in surface flux [e.g., *Schuh et al.*, 2010] and thus the concentration data. However, in cases when the diurnal cycle is fixed by the a priori flux estimate, any residuals caused by a mismatch between the "true" diurnal cycle seen by the tower and that prescribed by the fixed diurnal cycle of fluxes can be aliased onto other areas or regions in the inversion estimation.

3.4. Case 4: Isolating Influence of Near- and Far-Field Fluxes

[42] Case 4 examines the influence of flux variability in the far field on the synthetic observation time series (section 2.3, Figure 3d). Atmospheric data are not equally sensitive to an entire ecoregion or all areas surrounding the tower. Instead, observations tend to be most strongly influenced by sources and sinks within the immediate proximity of the towers [*Gerbig et al.*, 2009]. Case 4 examines the extent to which the results in Case 3 are impacted by differences in fluxes in areas farther away from the tower (e.g., outside of the tower's near field). The ability of the tower to detect flux differences in the far field depends on the tower's height, location, and the weather patterns or atmospheric transport, and will have a strong impact on whether a tower's observations can be used to evaluate flux distributions beyond its near field. Thus, Case 4 examines the spatial range seen by the towers, and therefore whether the atmospheric observations can be used to evaluate fluxes across a large domain, or only those areas very close to tower locations.

[43] Overall, much of the strong summertime differences among the signals in Case 3 appear to originate from flux variability (both spatial and temporal) within the near field (Figure 6). Although differences among signals still remain at many of the towers in Case 4, fewer pairs of signals have statistically significant differences. Thus, when near-field variability is removed in the TBM fluxes, the resultant signals are much less statistically significantly different. This result is consistent with other studies that examined the influence of near- and far-field fluxes on observations [e.g., *Gerbig et al.*, 2009]. However, at towers with larger footprints (e.g., LEF, CDL, and FRD), differences among fluxes in the far field do appear to translate into significant differences among the synthetic CO_2 signals. These towers are also located near regions with greater across-model variability in surface flux estimates between TBM (Figure 2b) and have the lowest model-data mismatch variances.

[44] As in the modified version of Case 3, the diurnal cycle was removed from the far-field fluxes to isolate the impact of fine-scale spatial and temporal differences in estimated fluxes from the TBMs. Contrary to the modified Case 3, however, the significance of the differences between concentration signals remained largely unchanged, indicating that most of the differences in signals observed in Case 4 are driven by differences in the spatial distribution of fluxes in the far field (Figure 6), rather than differences in

the far-field diurnal cycle as represented by each TBM. This is likely due to the fact that transported fluxes from regions farther away from tower locations are more likely to be well-mixed and diffuse, compared to near-field fluxes. Thus, differences in the diurnal cycle of fluxes in the far field are likely less detectable at the towers.

[45] The combined results from Cases 2 through 4 suggest that the observed differences in the CO_2 signal are driven primarily by differences in flux magnitude over large scales, and the diurnal cycle of fluxes within the near field of tower locations. Thus, the scale at which TBMs can be evaluated using atmospheric CO_2 observations may be limited to areas within the near field of tower observations, or aggregated net fluxes over large regions.

[46] The results from Case 4 are also important for inversions that use prior flux estimates derived from TBMs. Diurnal variability in surface fluxes within the near field of tower observations appear to have a significant impact on the high-frequency variations in the atmospheric data. Thus, in order to accurately account for the variability in the atmospheric data, the prior has to accurately resolve the flux at fine temporal scales over regions in close proximity to tower locations, or the inversion needs to be able to adjust the temporal (and spatial) variability of the prior.

4. Summary and Conclusions

[47] This study examined the ability of atmospheric measurements to detect differences in the 1° by 1° land-atmosphere carbon fluxes from four different TBMs. Motivated by the need to improve both forward and inversion models, this study examined three main applications using atmospheric data: (1) evaluating the relative merits of two or more TBMs; (2) validating a single TBM using a transport model and atmospheric concentrations; and (3) ability of the atmospheric network to provide sufficient information to accurately infer fine-scale flux variations in the context of atmospheric inversion.

[48] Using a Chi-square test of variance, the CO_2 signals generated using the different representations of surface flux were compared, and the results suggest that there is sufficient information in the atmospheric record to evaluate or validate at least some aspects of surface flux estimates from TBMs. Given the types of differences in flux represented by the four TBMs, atmospheric data may be most informative in evaluating aggregated TBM fluxes over large spatial scales (e.g., ecoregions), as well as in the improvement of how the diurnal cycle of fluxes is represented in TBMs, particularly in areas close to tower locations. As the density of the atmospheric network increases, more areas will fall within the near field of tower locations, and the monitoring network will thereby provide more information for model evaluation.

[49] The high sensitivity of the signals to slight differences in the diurnal cycle of fluxes stresses the importance of accurately accounting for small-scale temporal variability in fluxes in models, both in inversions and process-oriented TBMs. This becomes particularly important when inversions use prior flux estimates derived from TBMs. In order to accurately account for the high-frequency variations in the atmospheric data, the prior either needs to be very good at capturing the true flux variability at fine temporal scales, or the inversion needs to be able to adjust this variability.

Whereas studies focused on evaluating inversion setups have shown the potential impact of spatial aggregation errors on estimates [e.g., *Kaminski et al.*, 2001; *Schuh et al.*, 2009], the results of this study confirm that the time domain runs the same risk of aggregation errors [e.g., *Law et al.*, 2004; *Peylin et al.*, 2005; *Gourdji et al.*, 2010]. Results demonstrate that the impact of temporal aggregation may be equally important to the impact of aggregation in the spatial domain.

[50] The high sensitivity of concentration observations to the near field of tower locations further highlights the importance of potential aggregation errors in inversions. For example, to avoid spatial aggregation errors, some inversion approaches optimize fluxes at small spatial scales [e.g., *Gerbig et al.*, 2003a; *Rödenbeck et al.*, 2003; *Michalak et al.*, 2004; *Peylin et al.*, 2005; *Gourdji et al.*, 2008; *Schuh et al.*, 2010]. Such flexibility is important considering the results from Case 4, which suggest that measurements at many towers are highly influenced by small-scale flux variability in the near field. Unless an inversion can adjust small-scale variability, biased estimates may be expected at large scales. Furthermore, by properly accounting for the near-field, small-scale variability, inversions are likely to recover large-scale fluxes more accurately.

[51] One of the limitations of this study is that the results apply to individual towers, and do not quantify the compounding benefit of multiple towers. Thus, although two signals may not be statistically significantly different if looked at from the perspective of any one tower, the results might be different if the signals at multiple towers at considered together.

[52] Finally, the results from the significance testing are strongly influenced by the estimates of model-data mismatch variance used in this study. As the ability to model atmospheric transport improves (i.e., lower model-data mismatch variance), more subtle differences in surface flux representation will be detectable through the atmospheric data. Conversely, as TBMs improve, models may converge toward a similar solution, making the differences between models more difficult to detect.

References

Baker, D. F., et al. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003, *Global Biogeochem. Cycles*, *20*, GB1002, doi:10.1029/2004GB002439.

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. da Rocha (2008), Seasonal drought stress in the Amazon: Reconciling models and observations, *J. Geophys. Res.*, *113*, G00B01, doi:10.1029/2007JG000644.

Bousquet, P., P. Ciais, P. Peylin, M. Ramonet, and P. Monfray (1999), Inverse modeling of annual atmospheric CO$_2$ sources and sinks: 1. Method and control inversion, *J. Geophys. Res.*, *104*(D21), 26,161–26,178, doi:10.1029/1999JD900342.

Butler, M. P., K. J. Davis, A. S. Denning, and S. R. Kawa (2011), Using continental observations in global atmospheric inversions of CO$_2$: North American carbon sources and sinks, *Tellus, Ser. B*, *62*(5), 550–572, doi:10.1111/j.1600-0889.2010.00501.x.

Carouge, C., P. Bousquet, P. Peylin, P. J. Rayner, and P. Ciais (2010), What can we learn from European continuous atmospheric CO$_2$ measurements to quantify regional fluxes: Part 1. Potential of the 2001 network, *Atmos. Chem. Phys.*, *10*, 3107–3117, doi:10.5194/acp-10-3107-2010.

Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais (2006), On the assignment of prior errors in Bayesian inversions of CO$_2$ surface fluxes, *Geophys. Res. Lett.*, *33*, L13802, doi:10.1029/2006GL026496.

Cramer, W., D. W. Kicklighter, A. Bondeau, B. Moore, C. Churkina, B. Nemry, A. Ruimy, and A. L. Schloss (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, *5*, 1–15, doi:10.1046/j.1365-2486.1999.00009.x.

Dargaville, R. J., et al. (2002), Evaluation of terrestrial carbon cycle models with atmospheric CO$_2$ measurements: Results from transient simulations considering increasing CO$_2$, climate, and land-use effects, *Global Biogeochem. Cycles*, *16*(4), 1092, doi:10.1029/2001GB001426.

Denning, A. S., M. Nicholls, L. Prihodko, I. Baker, P. L. Vidale, K. Davis, and P. Bakwin (2003), Simulated variations in atmospheric CO$_2$ over a Wisconsin forest using a coupled ecosystem-atmosphere model, *Global Change Biol.*, *9*, 1241–1250, doi:10.1046/j.1365-2486.2003.00613.x.

Engelen, R. J., A. S. Denning, and K. R. Gurney (2002), On error estimation in atmospheric CO$_2$ inversions, *J. Geophys. Res.*, *107*(D22), 4635, doi:10.1029/2002JD002195.

Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9780511535741.

Geels, C., et al. (2007), Comparing atmospheric transport models for future regional inversions over Europe - Part 1: Mapping the atmospheric CO$_2$ signals, *Atmos. Chem. Phys.*, *7*, 3461–3479, doi:10.5194/acp-7-3461-2007.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003a), Toward constraining regional-scale fluxes of CO$_2$ with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, *108*(D24), 4756, doi:10.1029/2002JD003018.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003b), Toward constraining regional-scale fluxes of CO$_2$ with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, *108*(D24), 4757, doi:10.1029/2003JD003770.

Gerbig, C., J. C. Lin, J. W. Minger, and S. C. Wofsy (2006), What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, *Atmos. Chem. Phys.*, *6*, 539–554, doi:10.5194/acpd-5-9249-2005.

Gerbig, C., A. J. Dolman, and M. Heimann (2009), On observational and modelling strategies targeted at regional carbon exchange over continents, *Biogeosciences*, *6*(10), 1949–1959, doi:10.5194/bg-6-1949-2009.

Global Change Project (GCP) (2010), Regional Carbon Cycle Assessment and Processes (RECCAP) soft protocol, version 4, report, Canberra, ACT, Australia.

Gloor, M., P. Bakwin, D. Hurst, L. Lock, R. Draxler, and P. Tans (2001), What is the concentration footprint of a tall tower?, *J. Geophys. Res.*, *106*(D16), 17,831–17,840, doi:10.1029/2001JD900021.

Gourdji, S. M., K. L. Mueller, K. Schaefer, and A. M. Michalak (2008), Global monthly averaged CO$_2$ fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, *J. Geophys. Res.*, *113*, D21115, doi:10.1029/2007JD009733.

Gourdji, S. M., A. I. Hirsch, K. L. Mueller, A. E. Andrews, and A. M. Michalak (2010), Regional-scale geostatistical inverse modeling of North American CO$_2$ fluxes: A synthetic data study, *Atmos. Chem. Phys.*, *10*, 6151–6167, doi:10.5194/acp-10-6151-2010.

Gurney, K. R., et al. (2002), Towards robust regional estimates of CO$_2$ sources and sinks using atmospheric transport models, *Nature*, *415*(6872), 626–630, doi:10.1038/415626a.

Gurney, K. R., et al. (2003), TransCom 3 CO$_2$ inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, *Tellus, Ser. B*, *55*(2), 555–579, doi:10.1034/j.1600-0889.2003.00049.x.

Gurney, K. R., et al. (2004), Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, *Global Biogeochem. Cycles*, *18*, GB1010, doi:10.1029/2003GB002111.

Heimann, M., et al. (1998), Evaluation of terrestrial carbon cycle models through simulations of the seasonal cycle of atmospheric CO$_2$: First results of a model intercomparison study, *Global Biogeochem. Cycles*, *12*(1), 1–24, doi:10.1029/97GB01936.

Kaminski, T., and M. Heimann (2001), Inverse modeling of atmospheric carbon dioxide fluxes, *Science*, *294*(5541), 259, doi:10.1126/science.294.5541.259a.

Kaminski, T., M. Heimann, and R. Giering (1999), A coarse grid three-dimensional global inverse model of the atmospheric transport: 1. Adjoint model and Jacobian matrix, *J. Geophys. Res.*, *104*(D15), 18,535–18,553, doi:10.1029/1999JD900147.

Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*(D5), 4703–4715, doi:10.1029/2000JD900581.

Krinner, G., N. Viovy, N. de Noblet-Ducoudre, J. Ogee, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Lauvaux, T., M. Uliasz, C. Sarrat, F. Chevallier, P. Bousquet, C. Lac, K. J. Davis, P. Ciais, A. S. Denning, and P. J. Rayner (2008), Mesoscale inversion: First results from the CERES campaign with synthetic data, *Atmos. Chem. Phys.*, *8*, 3459–3471, doi:10.5194/acp-8-3459-2008.

Law, R. M., P. J. Rayner, and I. G. Enting (2002), Using high temporal frequency data for CO$_2$ inversions, *Global Biogeochem. Cycles*, *16*(4), 1053, doi:10.1029/2001GB001593.

Law, R. M., P. J. Rayner, and Y. P. Wang (2004), Inversion of diurnally varying synthetic CO$_2$: Network optimization for an Australian test case, *Global Biogeochem. Cycles*, *18*, GB1044, doi:10.1029/2003GB002136.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger (2003), A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *J. Geophys. Res.*, *108*(D16), 4493, doi:10.1029/2002JD003161.

Lin, J. C., C. Berbig, B. C. Daube, S. C. Wofsy, A. E. Andres, S. A. Vay, and B. E. Anderson (2004), An empirical analysis of the spatial variability of atmospheric CO$_2$: Implications for inverse analyses and space-borne sensors, *Geophys. Res. Lett.*, *31*, L23104, doi:10.1029/2004GL020957.

McGuire, A. D., et al. (2001), Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO$_2$, climate and land-use effects with four process-based ecosystem models, *Global Biogeochem. Cycles*, *15*(1), 183–206, doi:10.1029/2000GB001298.

Melillo, J. M., et al. (1995), Vegetation/Ecosystem Modeling and Analysis Project: Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate-change and CO$_2$ doubling, *Global Biogeochem. Cycles*, *9*(4), 407–437, doi:10.1029/95GB02746.

Michalak, A. M., L. Bruhwiler, and P. P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *J. Geophys. Res.*, *109*, D14109, doi:10.1029/2003JD004422.

Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *J. Geophys. Res.*, *110*, D24107, doi:10.1029/2005JD005970.

Nevison, C. D., N. M. Mahowald, S. C. Doney, I. D. Lima, G. R. Van der Werf, J. T. Randerson, D. F. Baker, P. Kasibhatla, and G. A. McKinley (2008), Contribution of ocean, fossil fuel, land biosphere, and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO$_2$, *J. Geophys. Res.*, *113*, G01010, doi:10.1029/2007JG000408.

Olsen, S. C., and J. T. Randerson (2004), Differences between surface and column atmospheric CO$_2$ and implications for carbon cycle research, *J. Geophys. Res.*, *109*, D02301, doi:10.1029/2003JD003968.

Olson, D. M., et al. (2001), Terrestrial ecoregions of the worlds: A new map of life on Earth, *BioScience*, *51*(11), 933–938, doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Peters, W., et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(48), 18,925–18,930, doi:10.1073/pnas.0708986104.

Peylin, P., P. Bousquet, and P. Ciais (2001), Inverse modeling of atmospheric carbon dioxide fluxes - Response, *Science*, *294*(5541), U1–U2.

Peylin, P., D. Baker, J. Sarmiento, P. Ciais, and P. Bousquet (2002), Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO$_2$ data, *J. Geophys. Res.*, *107*(D19), 4385, doi:10.1029/2001JD000857.

Peylin, P., P. J. Rayner, P. Bousquet, C. Carouge, F. Hourdin, P. Heinrich, P. Ciais, and AEROCARB contributors (2005), Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1. Inverse methodology, *Atmos. Chem. Phys.*, *5*, 3173–3186, doi:10.5194/acp-5-3173-2005.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993), Terrestrial ecosystem production: A process model-based on global satellite and surface data, *Global Biogeochem. Cycles*, *7*(4), 811–841, doi:10.1029/93GB02725.

Randerson, J. T., M. V. Thompson, T. J. Conway, I. Y. Fung, and C. B. Field (1997), The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cycles*, *11*(4), 535–560, doi:10.1029/97GB02268.

Rodell, M., et al. (2004), The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, *85*(3), 381–394, doi:10.1175/BAMS-85-3-381.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964, doi:10.5194/acp-3-1919-2003.

Schimel, D. S., et al. (1997), Continental scale variability in ecosystem processes: Models, data, and the role of disturbance, *Ecol. Monogr.*, *67*(2), 251–271, doi:10.1890/0012-9615(1997)067[0251:CSVIEP]2.0.CO;2.

Schuh, A. E., A. S. Denning, M. Uliasz, and K. D. Corbin (2009), Seeing the forest through the trees: Recovering large scale carbon flux biases in the midst of small-scale variability, *J. Geophys. Res.*, *114*, G03007, doi:10.1029/2008JG000842.

Schuh, A. E., A. S. Denning, K. D. Corbin, I. T. Baker, M. Uliasz, N. Parazoo, A. E. Andrews, and D. E. J. Worthy (2010), A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, *7*(5), 1625–1644, doi:10.5194/bg-7-1625-2010.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A simple biosphere model (SiB) for use within general-circulation models, *J. Atmos. Sci.*, *43*(6), 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the advanced research WRF version 2, *NCAR Tech. Note NCAR/TN–468+STR*, 88 pp, NCAR, Boulder, Colo.

van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and A. F. Arellano (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, *6*, 3423–3441, doi:10.5194/acp-6-3423-2006.

Wang, J. W., A. S. Denning, L. X. Lu, I. T. Baker, K. D. Corbin, and K. J. Davis (2007), Observations and simulations of synoptic, regional, and local variations in atmospheric CO_2, *J. Geophys. Res.*, *112*, D04108, doi:10.1029/2006jd007410.

Waring, R. H., and S. W. Running (2007), *Forest Ecosystems: Analysis at Multiple Scales*, 3rd ed., Elsevier, New York.

Zeng, N. (2003), Glacial-interglacial atmospheric CO_2 change—The glacial burial hypothesis, *Adv. Atmos. Sci.*, *20*(5), 677–693, doi:10.1007/BF02915395.

Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO_2 variability, *Global Biogeochem. Cycles*, *19*, GB1016, doi:10.1029/2004GB002273.

S. M. Gourdji, D. N. Huntzinger, A. M. Michalak, and K. L. Mueller, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA. (dnhuntzi@umich.edu)