North American CO\textsubscript{2} exchange: intercomparison of modeled estimates with results from a fine-scale atmospheric inversion

S. M. Gourdji\textsuperscript{1,8}, K. L. Mueller\textsuperscript{1}, V. Yadav\textsuperscript{1,7}, D. N. Huntzinger\textsuperscript{1,9}, A. E. Andrews\textsuperscript{2}, M. Trudeau\textsuperscript{2}, G. Petron\textsuperscript{2}, T. Nehrkorn\textsuperscript{3}, J. Eluszkiewicz\textsuperscript{3}, J. Henderson\textsuperscript{3}, D. Wen\textsuperscript{4}, J. Lin\textsuperscript{4}, M. Fischer\textsuperscript{5}, C. Sweeney\textsuperscript{2}, and A. M. Michalak\textsuperscript{1,6,7}

\textsuperscript{1}Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA
\textsuperscript{2}Global Monitoring Division, Earth System Research Laboratory, National Oceanic & Atmospheric Administration, Boulder, CO 80305, USA
\textsuperscript{3}Atmospheric and Environmental Research Inc., Lexington, MA 02421, USA
\textsuperscript{4}Department of Earth & Environmental Sciences, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
\textsuperscript{5}Lawrence Berkeley National Lab, Berkeley, CA 94720, USA
Department of Atmospheric, Oceanic & Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA

7Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA

8Department of Environmental Earth System Science and Program on Food Security & the Environment, Stanford University, Stanford, CA 94305, USA

9School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff AZ, USA

Received: 7 June 2011 – Accepted: 17 June 2011 – Published: 11 July 2011

Correspondence to: S. M. Gourdji (sgourdji@umich.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Robust estimates of regional-scale terrestrial CO$_2$ exchange are needed to support carbon management policies and to improve the predictive ability of models representing carbon-climate feedbacks. Large discrepancies remain, however, both among and between CO$_2$ flux estimates from atmospheric inverse models and terrestrial biosphere models. Improved atmospheric inverse models that provide robust estimates at sufficiently fine spatial scales could prove especially useful for monitoring efforts, while also serving as a validation tool for process-based assumptions in terrestrial biosphere models. A growing network of continental sites collecting continuous CO$_2$ measurements provides the information needed to drive such models. This study presents results from a regional geostatistical inversion over North America for 2004, taking advantage of continuous data from the nine sites operational in that year, as well as available flask and aircraft observations. The approach does not require explicit prior flux estimates, resolves fluxes at finer spatiotemporal scales than previous North American inversion studies, and uses a Lagrangian transport model coupled with high-resolution winds (i.e. WRF-STILT) to resolve near-field influences around measurement locations. The estimated fluxes are used in an inter-comparison with other inversion studies and a suite of terrestrial biosphere model estimates collected through the North American Carbon Program Regional and Continental Interim Synthesis. Differences among inversions are found to be smallest in areas of the continent best-constrained by the atmospheric data, pointing to the value of an expanded measurement network. Aggregation errors in previous coarser-scale inversion studies are likely to explain a portion of the remaining spread. The spatial patterns from a geostatistical inversion that includes auxiliary environmental variables from the North American Regional Reanalysis were similar to those from the median of the biospheric model estimates during the growing season, but diverged more strongly in the dormant season. This could be due to a lack of sensitivity in the inversion during the dormant season, but may also point to a lack of skill in the biospheric models outside of the growing season, particularly in agricultural
areas. For the annual continental budget, the boundary conditions used as an input into the inversions were seen to have a substantial impact on the estimated net flux, with a difference of \( \sim 0.8 \text{ PgC yr}^{-1} \) associated with results using two different plausible sets of boundary conditions.

1 Introduction

Carbon cycle scientists are increasingly called upon to provide information in support of efforts to monitor anthropogenic \( \text{CO}_2 \) emissions, and to provide predictions of future changes to the carbon cycle within the context of a changing climate and land-use choices (CCSP, 2007). Atmospheric inverse models can contribute to these goals by taking advantage of the information contained in atmospheric \( \text{CO}_2 \) mixing ratio measurements regarding upwind surface \( \text{CO}_2 \) exchange. Using these measurements, together with an atmospheric transport model and within a robust statistical framework (e.g. Enting, 2002), inverse models are used to infer the spatiotemporal distribution and magnitude of surface \( \text{CO}_2 \) fluxes. In addition, inverse-modeling derived \( \text{CO}_2 \) flux estimates that reflect the atmospheric data constraint are potentially useful for validating the process-based formulations of terrestrial ecosystem models. In fact, atmospheric measurements of \( \text{CO}_2 \) provide one of the key means of evaluating mechanistic models (e.g. Randerson et al., 2009; Cadule et al., 2010), given the lack of direct flux observations at regional scales.

Global inverse models have been used to infer continental-scale fluxes (e.g. Gurney et al., 2002; Rödenbeck et al., 2003; Baker et al., 2006; Mueller et al., 2008; Gourdji et al., 2008) by coupling global atmospheric transport models to atmospheric \( \text{CO}_2 \) observations, most typically weekly flask samples from approximately 80 locations worldwide in the NOAA-ESRL Cooperative Air Sampling Network (Tans and Conway, 2005). Because the sites in this measurement network were selected to sample well-mixed air from marine boundary layer and other locations remote from areas with strong and highly variable \( \text{CO}_2 \) fluxes, these global inversions are not able to reliably constrain fluxes at sub-continental scales.
Inversions that can take advantage of spatial and temporal atmospheric CO$_2$ gradients measured in areas with high flux variability provide the potential to resolve finer sub-continental scale fluxes. Resolving fluxes at these finer spatial scales is essential for supporting carbon management efforts at the level of countries, states and provinces, as well as for validating mechanistic models of the carbon cycle. Towards this end, an expanding in situ measurement network in North America and Europe (e.g. NOAA-ESRL, 2011; CGGMN, 2011; CEAD, 2011) is providing continuous measurements of atmospheric CO$_2$ mixing ratios. However, optimally extracting the flux signal from these data presents a challenging problem due to the combined influence on atmospheric CO$_2$ mixing ratios of the diurnal cycle of the terrestrial biosphere, heterogeneous land cover, point source fossil fuel emissions and complex atmospheric transport (Bakwin et al., 1998). Therefore, simultaneous improvements in inversion setups (e.g. Law et al., 2002; Schuh et al., 2009; Gourdji et al., 2010) and in the quality of atmospheric transport models (e.g. Geels et al., 2007) have been necessary for appropriately making use of these new data-streams within inversions.

Regional inversions that estimate fluxes for a limited domain have also emerged as a means to take advantage of continuous data collected at continental sites. By limiting the domain size, fluxes can be estimated at relatively fine spatial scales (e.g. $1^{\circ} \times 1^{\circ}$), thereby reducing aggregation errors (e.g. Kaminski et al., 2001; Engelen et al., 2002) associated with estimating coarse-scale fluxes using highly variable CO$_2$ measurement data, while simultaneously keeping the computational cost of inversions manageable. In addition, with a limited domain, it is possible to take advantage of high-resolution meteorological information and Lagrangian transport models that can resolve atmospheric dynamics in the near-field of measurement locations (e.g. Lin et al., 2003; Gerbig et al., 2008). However, regional inversions bring other complications relative to global inversions, including the need to specify atmospheric CO$_2$ boundary conditions for the region of interest, which have been shown to have a large influence on the resulting flux estimates (Peylin et al., 2005; Göckede et al., 2010b).
While continuous data-streams hold promise for estimating CO₂ fluxes at finer spatial scales, flux estimates from regional inversions are difficult to evaluate given the complexity associated with inversion setup and input choices (e.g. Ciais et al., 2010). For example, observed differences in estimates across inverse modeling studies could be due to a number of factors: atmospheric transport, flux resolution, boundary conditions, errors in the bottom-up models used as priors, treatment of fossil fuels, and the temporal resolution of observations, among other possible causes.

A similarly wide, if not wider, spread has also been seen in inter-comparisons of bottom-up, mechanistically modeled estimates of biospheric CO₂ flux (Huntzinger et al., 2011c; Hayes et al., 2011), creating a challenge in reconciling top-down and bottom-up estimates. Formal model inter-comparison studies that systematically evaluate the influence of model formulation, processes and driver data on flux estimates have begun to emerge as a means to reconcile observed differences between terrestrial biospheric models (e.g. the Multi-Scale Synthesis and Terrestrial Biospheric Model Intercomparison Project, or MsTMIP, 2011). Similar efforts are also underway to develop a formal framework for biospheric model-data intercomparison (the International Land Model Benchmarking Project, or iLAMB, 2011). Finally, given that atmospheric CO₂ mixing ratio measurements sampled in space and time provide the most direct evidence of upwind surface CO₂ fluxes, a reconciliation of top-down flux estimates across inversion studies could be especially useful for validating biospheric model estimates and their process-based assumptions.

As a step towards reconciling CO₂ flux estimates from regional inversions, this study presents a regional grid-scale geostatistical inversion (Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008, 2010) for North America in 2004, using the sampling network of 9 towers collecting continuous CO₂ measurements at that time, as well as available surface flask and aircraft data. Results are compared at various spatial and temporal scales to estimates from other inversion studies over North America to help illuminate potential causes of their differences and associated strengths and weaknesses of various approaches. The geostatistical approach implemented here uses
the optimized setup from Gourdji et al. (2010), which resolves fluxes at finer spatial and temporal scales than other published inversion studies for the same domain (e.g. Peters et al., 2007; Deng et al., 2007; Schuh et al., 2010; Butler et al., 2010). Also, by eliminating the requirement for explicit prior flux estimates and optimizing covariance parameters directly with the atmospheric data, the inversions presented here reduce potential biases associated with these setup choices. Geostatistical inversion-derived flux estimates are also used to interpret the wide spread seen across a collection of biospheric models participating in the North American Carbon Program (NACP) Regional and Continental Interim Synthesis study (RCIS; Huntzinger et al., 2011c).

2 Data and methods

The inversions presented here use a setup optimized as part of a synthetic data study with a similar tower configuration for June 2004 (i.e. Gourdji et al., 2010). This setup was also evaluated using synthetic data for a full year to guide the presentation and interpretation of the results presented in this paper. An overview of the setup and methods implemented in this study is provided below, with additional details provided in the Supplement.

2.1 Flux domain and resolution

Following Gourdji et al. (2010), fluxes are estimated in this study at a $1^\circ \times 1^\circ$, 3-hourly resolution. It should be noted that flux estimates are unlikely to be trustworthy at this fine resolution due to atmospheric mixing and the sparse network; however, biases due to estimating fluxes directly at large scales, termed aggregation errors (Kaminski et al., 2001; Engelen et al., 2002), are minimized by estimating fluxes first at fine scales and then post-aggregating. Spatial aggregation errors have been shown to be particularly problematic when making use of continental, continuous CO$_2$ measurements in inversions (Gerbig et al., 2003b; Schuh et al., 2009). Temporal aggregation errors are also
a concern when the shape of the diurnal cycle is fixed from biospheric models (Gourdji et al., 2010; Huntzinger et al., 2011b), rather than estimated directly as in the current study.

The North American domain spans from 10° N to 70° N and 50° W to 170° W, yielding 2635 land grid-cells (Fig. 1). Along with the 3-hourly timescale, this results in approximately 8 million estimated fluxes for the year (2635 regions × 366 days × 8 flux periods per day). Resolving fluxes for a 4-day averaged diurnal cycle temporal resolution (Gourdji et al., 2010) was explored, but the penalty in terms of the quality of the flux estimates was found to be unacceptable in this study.

2.2 Geostatistical inversions

Geostatistical inverse modeling (GIM, e.g. Hoeksema and Kitanidis, 1984; Zimmerman et al., 1998) has been used in atmospheric applications to identify trace gas sources and sinks (e.g. Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008, 2010). Geostatistical CO$_2$ inversions are Bayesian, but differ from synthesis Bayesian inversions (e.g. Baker et al., 2006; Peters et al., 2007; Butler et al., 2010) in two key ways.

First, GIM does not rely on a set of explicit prior flux estimates derived from biospheric models, fossil fuel inventories, fire emission estimates, and/or ocean flux estimates as in synthesis Bayesian inversion studies (e.g. Peters et al., 2007; Butler et al., 2010). Instead, the prior is defined using a set of predictor variables, in a manner analogous to multi-linear regression. The coefficients associated with these variables are estimated simultaneously with the fluxes using the atmospheric data constraint. With a very simple matrix of flux covariates (representing one or more mean fluxes in time and space), a geostatistical inversion can provide a completely independent comparison to biospheric models. Alternatively, datasets representative of different processes driving carbon exchange, defined at the grid-scale, can be included, e.g. fossil fuel inventories, remotely sensed vegetative indices, climatological model output, or even biospheric model estimates of CO$_2$ flux. Overall, by introducing more flexibility into the prior, GIM yields flux estimates that are less subject to biases in the biospheric models.
used as prior flux estimates in other approaches. In addition, the inferred coefficients on the predictor variables can provide insights into flux drivers at the resolution of the inversion.

Secondly, GIM is well suited for estimating fluxes directly at fine scales, by incorporating a priori information on the spatial and/or temporal covariance between estimated fluxes (or flux residuals). The parameters describing this covariance structure are also optimized directly using the atmospheric measurements, as described further in Sect. 2.4. While recent synthesis Bayesian inversion studies also estimate fluxes at relatively fine spatial scales to avoid aggregation errors, and rely on spatial covariance assumptions to provide an additional constraint (e.g. Rödenbeck et al., 2003; Carouge et al., 2010a, b; Chevallier et al., 2010; Schuh et al., 2010), the covariance parameters have been based on indirect information, including analyses of variability in biospheric models, or sensitivity testing to assess their impact on flux estimates.

The objective function $L_{s,\beta}$ for GIM is:

$$L_{s,\beta} = \frac{1}{2} (z - Hs)^T R^{-1} (z - Hs) + \frac{1}{2} (s - X\beta)^T Q^{-1} (s - X\beta)$$  \hspace{1cm} (1)$$

where the vector $z$ ($n \times 1$) represents the atmospheric CO$_2$ measurements (ppm), and $s$ ($m \times 1$) is the vector of surface fluxes (µmol (m$^{-2}$ s$^{-1}$)). $H$ ($n \times m$) describes the sensitivity of CO$_2$ measurements to fluxes, as quantified from an atmospheric transport model, with units of ppm/(µmol (m$^{-2}$ s$^{-1}$)), and $Hs$ therefore represents a vector of modeled CO$_2$ observations. $X$ is a known ($m \times p$) matrix containing the pre-selected flux covariates, $\beta$ are ($p \times 1$) unknown regression coefficients on these covariates, and $X\beta$ is the component of the flux variability that can be explained by the covariates, a.k.a. the “trend” This objective function is almost identical to that for a synthesis Bayesian inversion, except that here, the $X\beta$ term replaces the explicit prior flux estimates, and the unknown fluxes ($s$) and regression coefficients ($\beta$) are optimized simultaneously.

The two covariance matrices in the objective function, $R$ ($n \times n$) and $Q$ ($m \times m$), balance the relative weight of the atmospheric data and the trend in estimating fluxes. $R$ is the model-data mismatch covariance matrix, describing the expected magnitude of
discrepancies between observed (z) and modeled (Hs) CO₂ mixing ratios (due to measurement, transport, representation, and aggregation errors). Q (m × m) is the a priori flux covariance matrix, characterizing how flux deviations from the model of the trend (i.e. s – Xβ) are correlated in time and space. In GIM, the diagonal of Q represents the expected variance of flux residuals from the trend, while in a synthesis Bayesian inversion, this would represent the uncertainties associated with the explicit prior flux estimates.

Altogether the best estimates of flux (ŝ) are a composite of the estimated trend, or deterministic, component (X̂β), and a spatially correlated stochastic component, which describes the portion of the flux signal that is seen through the atmospheric data, but which cannot be explained by the linear trend X̂β. Fluxes (ŝ) and regression coefficients (̂β) are estimated by minimizing the GIM objective function (Eq. 1) solving the linear system of equations shown in Supplement A (Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008).

2.3 Atmospheric data, boundary conditions, and transport

2.3.1 Atmospheric CO₂ mixing ratio measurements

This study uses continuous, high-precision, well-calibrated CO₂ mixing ratio measurements from 9 observational locations unevenly spaced across the North American continent in 2004 (Fig. 1). These include two tall towers with a height of 457 m (Moody, Texas) and 396 m (Park Falls, Wisconsin), two coastal towers less than 25 m in height (Sable Island, Nova Scotia and Barrow, Alaska), and five other inland, continental towers ranging in height from 30 to 107 m (Norman, Oklahoma, Harvard Forest, Massachusetts, Argyle, Maine, Fraserdale, Ontario, and Candle Lake, Saskatchewan). In addition, to maximize the atmospheric data constraint, all available flask and aircraft measurements are also included with the exception of flask samples at coincident tower locations and some coastal sites where the atmospheric transport model was deemed unreliable. Supplement B provides additional details on CO₂ data processing and filtering.
Please note that the North American measurement network has since expanded to more than 40 sites collecting continuous CO\textsubscript{2} mixing ratio measurements (Mueller et al., 2011). However, many new sites are in complex terrain, urban areas, or are very short, and the optimal use of this data in inversions remains a research topic in itself (Manning, 2011). Therefore, while this study was conducted for a relatively data-poor year before the expansion of the atmospheric monitoring network (2004), the results shown here provide a baseline for improving inversions taking advantage of a more data-rich environment.

2.3.2 Continental boundary conditions

Regional inversions necessitate the use of boundary conditions that represent the CO\textsubscript{2} concentrations of air flowing into the domain of interest (i.e. the North American land mass here). The impact of these boundary CO\textsubscript{2} mixing ratios on the observations used in the inversion must be pre-subtracted before inferring CO\textsubscript{2} fluxes over North America. Two plausible sets of CO\textsubscript{2} boundary conditions are used in this study: one optimized as part of the CarbonTracker global CO\textsubscript{2} data assimilation system (Peters et al., 2007), and the other derived more empirically from marine boundary layer and aircraft observations taken from the GLOBALVIEW-CO\textsubscript{2} (2010) database. Supplement B provides additional details on these datasets.

The majority of presented results use the empirical boundary conditions, but results are presented with both sets of boundary conditions where these were seen to have a large impact on the conclusions.

2.3.3 Atmospheric transport model

Surface influence functions (“footprints”, or adjoint sensitivities) express the sensitivity of individual CO\textsubscript{2} measurements at specific points in space and time to surface fluxes in the upwind source regions (in units of ppm/(\(\mu\text{mol m}^{-2} \text{s}^{-1}\))). The Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003), driven by
meteorological fields from the Weather Research and Forecasting (WRF) model (Skamarock and Klemp, 2008), customized for STILT (Nehrkorn et al., 2010), was used to derive these footprints. The WRF-STILT framework is well-suited for this application, because: (1) WRF meteorology is available at higher resolution than that used in most global models, and therefore has the potential to be more realistic (Mass et al., 2002); (2) the Lagrangian approach minimizes numerical diffusion present in Eulerian models (Odman, 1997) and is thus better able to represent plumes in the near-field of the measurement locations (Lin et al., 2003; Wen et al., 2011); (3) the WRF-STILT coupling has been specifically designed to achieve good mass conservation characteristics by using time-averaged winds from WRF within STILT (Nehrkorn et al., 2010); and (4) the Lagrangian approach offers the most efficient way to compute the grid-scale footprints, by running transport backwards in time (Lin et al., 2003). The computational aspects of the footprint calculations are described in more detail in Supplement C.

In addition to constraining fluxes in the inversion, the footprints are also used to assess which portions of the continent are constrained by a specific set of measurements. Given the limited network collecting continuous CO₂ measurements in 2004, not all portions of North America are equally well-constrained, as can be seen in the annual average footprint shown in Fig. 2a. Not surprisingly, the best-constrained areas are near the measurement locations in the Central and Eastern continental United States (US) and a large portion of Canada. The areas with a partial constraint are in the Western and Southeastern continental US, while the most under-constrained areas are in Mexico and Central America, Northwest Canada and Alaska.

Figure 2b shows three biomes across North America with similar land-cover and/or climatic characteristics (modified from Olson, 2001) that are relatively well-constrained by the measurement network. These three biomes are used for spatial aggregation and comparison of flux estimates across all inversion studies and biospheric models. (Please note that these biomes are not the native spatial resolution for any of the inversion studies included in the flux estimate comparison.) The Temperate
Grass/Savannah/Shrub, or the agricultural areas of the continent, is the biome best-constrained by the measurement network, and the Boreal Forest is relatively well-constrained, particularly in the central portion. The Eastern Temperate Forests are well-constrained in the Eastern and Midwestern US and Southern Canada, but lack sensitivity in the Southeast US.

2.4 Covariance matrix structure and parameter optimization

The $Q$ matrix describes the spatial correlation structure of flux residuals that cannot be explained by the predictor variables included in the matrix of flux covariates ($X$). Two parameters, the variance ($\sigma^2$), and the spatial correlation length parameter ($l$), are used to populate the $Q$ matrix assuming an isotropic exponential decay, which varies as a function of separation distance ($h$):

$$Q_{ij}(h_{ij}|\sigma^2,l) = \sigma^2 \exp \left( -\frac{h_{ij}}{l} \right)$$

The practical correlation length is approximately $3l$, beyond which $\sigma^2$ represents the expected variance of independent flux residuals. These spatial covariance parameters are also allowed to vary at a monthly timescale, given that these parameters have been found to have a strong seasonal cycle (Huntzinger et al., 2011a) that should be appropriately accounted for in order to yield realistic flux estimates from the inversion.

A priori temporal covariance assumptions in $Q$ were additionally considered, as in Gourdji et al. (2010), but were found to yield unrealistic extrapolated flux patterns in this study particularly in times of the year with rapid change (e.g. leaf-out). However, a priori temporal covariance assumptions were also shown in Gourdji et al. (2010) to be necessary for recovering realistic a posteriori uncertainties at temporally aggregated scales. Similarly, the estimated uncertainties from this study without temporal covariance assumptions were seen to be artificially low when post-aggregated to monthly and annual timescales. Therefore, the analytical uncertainties that correspond with the
presented fluxes are not shown in this study. The recovery of realistic confidence intervals from inversions, particularly at the annual timescale, remains an area of active research (e.g. Peters et al., 2010a).

The model-data mismatch covariance matrix \( (R) \) describes how well the “true” fluxes should be able to match the recorded measurements, given errors associated with atmospheric transport, the measurement instruments, and the coarse grid of the inversion and transport model relative to the scales of variability in the true fluxes (e.g. Kaminski et al., 2001; Engelen et al., 2002). In the current study, the model-data mismatch matrix \( (R) \) remains diagonal as in previous inversion studies, although as in Gourdji et al. (2010), a different value is optimized for each tower. This allows the inversion to de-weight continuous measurement locations where the transport model may be of poorer quality or where local variability in the measurement data is difficult to resolve. Separate model-data mismatch values are optimized for flask and aircraft data, although these values represent an average across sites. All optimized model-data mismatch values are also allowed to change monthly in order to account for seasonal variations in the quality of the transport model and inversion setup.

The covariance parameters used to construct the \( R \) and \( Q \) covariance matrices are estimated simultaneously using Restricted Maximum Likelihood (RML; e.g. Kitanidis, 1995), Michalak et al. (2004) with the atmospheric measurements, as described further in Gourdji et al. (2010). Supplement D provides additional details on the covariance parameters used for the inversions.

2.5 Auxiliary variables and variable selection

2.5.1 Biospheric flux covariates

Geostatistical inversions can estimate fluxes with or without the use of flux covariates. As in multi-linear regression, adding all possible covariates can help to improve the fit of the model to the data, although at the risk of introducing spurious relationships that could potentially bias flux estimates in under-constrained regions and time periods. For
this study, the Bayes Information Criterion (BIC) (Schwarz, 1978) variable selection technique is implemented to choose covariates that optimally explain the biospheric flux signal in the atmospheric data. The BIC was combined with the Branch-and-Bound algorithm (Yadav et al., 2011) in order to reduce computational expense. Supplement E presents the details of the implementation of these techniques.

Two main sets of inversion results are presented in this paper. For the first inversion, we include only a fossil fuel inventory dataset in $X$ with no additional information regarding biospheric processes. The fossil fuel inventory is included to help localize the spatial patterns of the fossil fuel emissions, which have a very different spatial structure from the biospheric fluxes. This setup is referred to as the “Simple” inversion in this paper, given that the atmospheric observations provide the only data constraint on biospheric fluxes. For the second inversion, we additionally incorporate auxiliary variables from the North American Regional Reanalysis (NARR; Mesinger et al., 2006). This case is termed the “NARR” inversion.

The NARR data products are a state-of-the-science meteorological reanalysis for North America, and have significantly improved the representation of the hydrological cycle relative to previous datasets (Bukovsky and Karoly, 2007). These variables thus have the potential to significantly improve CO$_2$ flux estimates. Overall, eleven datasets, defined for all flux locations and time periods, were considered for inclusion in the inversion, as well as two derived precipitation variables (average precipitation over the previous 16 and 30 day intervals). A subset of these 13 variables (Table 2) was then selected for the flux covariate matrix ($X$) using the combined BIC and Branch-and-Bound algorithm.

The superset of NARR variables shown in Table 2 primarily relate to the hydrological cycle, although some of the variables such as shortwave radiation, evapotranspiration and canopy conductance have a direct physiological relationship with photosynthetic CO$_2$ fluxes. Vegetative indices from remote-sensing datasets, such as Leaf Area Index or Fraction of Photosynthetically Active Radiation from the MODIS instruments (e.g. Yang et al., 2006), could also provide useful information regarding the seasonal
cycle and spatial distribution of CO$_2$ flux (e.g. Gourdji et al., 2008). Because such data are only available at a weekly timescale, they were not included here. Also, the evapotranspiration and canopy conductance variables from NARR are calculated with the Noah Land Surface Model (Ek et al., 2003), and therefore implicitly include the Normalized Difference Vegetation Index from the AVHRR instrument, an alternative to vegetation indices from the MODIS sensors.

### 2.5.2 Fossil fuel inventory

The fossil fuel inventory used in this study is a merged data product providing full coverage for the continent. In the continental United States, we take advantage of diurnally and seasonally varying estimates from version 1.4 of the Vulcan database (Gurney et al., 2009). These estimates for 2002 are scaled up to 2004 total emissions for the region, but without any changes in the underlying spatial and temporal patterns. In Central America, Mexico and Canada, emission estimates are taken from a monthly-varying dataset specifically for 2004, which merges information from British Petroleum fuel statistics, remotely-sensed night lights, and the existing Carbon Dioxide Information Analysis Center (CDIAC) fossil fuel emission inventory (Oda and Maksyutov, 2011). This inventory is used as a covariate in the Simple inversion.

In the NARR inversion, the impact of fossil fuels is removed from the atmospheric observations a priori, by multiplying the inventory dataset by the footprints and subtracting the resulting signal from the atmospheric measurements. By pre-subtracting the fossil fuel influence from the measurements, we reduce potential covariance in the inferred regression coefficients between the fossil fuel inventory and biospheric datasets due to covariance in the underlying processes, e.g. re-growing forests and high emissions in the Eastern continental United States, or reduced populations and industrial activity in arid and snow-covered areas. Such covariance would confound flux interpretation by making it more difficult to separate the biospheric and fossil fuel signals a posteriori. Regardless of whether the fossil fuel influence is subtracted from the measurements a priori, or from the flux estimates a posteriori, errors in the fossil fuel inventories will
become aliased onto the inferred biospheric fluxes, although these errors are thought
to be small relative to the uncertainties in terrestrial biospheric models (Marland et al.,
2009). Also, results from the Simple inversion (see Sect. 3.1.1) showed a recovered
regression coefficient on the fossil fuel inventory near one, implying that the inventory
dataset used in this study was a reasonable approximation for the “true” emissions
pattern and magnitude.

2.6 Evaluation of inferred fluxes and comparison to estimates from other
models

Because CO₂ fluxes cannot be measured directly at scales compatible with inversion-
derived flux estimates, validating inversions remains a challenge. Some studies have
taken the approach of excluding some atmospheric measurements from use in the
inversion (e.g. aircraft measurements), and then comparing modeled concentrations
resulting from inferred fluxes to these excluded observations to evaluate the inversion
(e.g. Peters et al., 2007; Chevallier et al., 2010). However, given the limited amount of
atmospheric CO₂ measurements in 2004, all available data-streams were used here
to help improve flux estimates and fill in regions of the continent otherwise under-
constrained by the nine tower network. Also, by comparing modeled and actual CO₂
concentrations, it is difficult to interpret the relative impact of biases in the estimated
flux distribution vs. errors in the transport model, limiting the power of such a tech-
nique. Other studies have compared inversion output to flux tower data (e.g. Schuh et
al., 2010), although the spatial scale mismatch also makes these comparisons incon-
clusive.

In addition to synthetic data inversions, as in Gourdji et al. (2010), which provide
a baseline “best case” scenario for the performance of the inversion, this study relies
on an inter-comparison with other top-down and bottom-up model output to assess the
robustness and limitations of the presented flux estimates. After isolating the biospheric
component in the a posteriori total flux estimates from both inversions, GIM results are
compared with those from previous inversion studies (CarbonTracker, 2009, or Peters
et al., 2007; Schuh et al., 2010; Butler et al., 2010) for the same domain, and also with a suite of bottom-up flux estimates from 16 terrestrial biospheric models that participated in the NACP RCIS (Huntzinger et al., 2011c). Table 1 compares the features of the GIM inversions conducted here, and the synthesis Bayesian inversions included in the inter-comparison. In addition to the details presented in Table 1, each inversion has its own particularities in terms of data processing and selection, identification of covariance parameters and numerical approaches for implementing the inversion scheme. Despite these differences, the atmospheric measurements included in the inversions for this year remains similar across studies, and the well-constrained areas of the continent should therefore also be similar. The biospheric models from the NACP RCIS included in the inter-comparison are also described in more detail in Huntzinger et al. (2011c).

Model estimates are compared across studies at monthly and annual timescales, and at grid and aggregated spatial scales. The analysis of grid-scale patterns helps to visually assess model output, while the comparison at aggregated spatial scales helps to evaluate the biospheric models with the atmospheric data constraint, particularly for the monthly seasonal cycle in well-constrained areas of the continent. Net annual flux estimates from all models are less reliable, given that they represent a small residual on a strongly-varying seasonal cycle. However, results are still presented at this longer timescale due to the larger scientific and societal interest in understanding the total carbon budget for North America, and the locations of net sources and sinks within the continent.
3 Results and discussion

3.1 Auxiliary variables and regression coefficients

3.1.1 Fossil fuel emissions inventory

If the fossil fuel emissions inventory and atmospheric transport model used in this study were perfect, we would expect the regression coefficient associated with the inventory to be approximately one. Values other than one could imply problems with the inversion setup, systematic transport model errors, a lack of atmospheric constraint, and/or errors in the spatiotemporal patterns and magnitudes of emissions in the inventory dataset.

Using the Simple inversion setup with the empirical boundary conditions yielded a $\hat{\beta}$ on the inventory of 1.00, whereas that for the Simple inversions using the Carbon-Tracker boundary conditions was 0.98 (with an uncertainty of $\sigma_{\hat{\beta}} = 0.05$ for both cases). This result provides support both for the quality of the inventory as well as for the inversion setup and components used here. In addition, this result makes it easy to separate the biospheric and fossil fuel contributions to the total flux a posteriori. The possibility that the recovered regression coefficient of one is due to multiple errors that cancel out (e.g. underestimated emissions in the inventory dataset in the near vicinity of the towers could compensate for a less than perfect correlation with the spatiotemporal patterns in the inventory) cannot, however, be eliminated.

3.1.2 NARR variables

Introducing covariates associated with the biospheric signal into the $X$ matrix gives an opportunity to identify significant flux drivers, and learn from their inferred relationships to CO$_2$ flux. The regression coefficients for the variables in this study represent an average relationship over the entire continent and year, and the selection of variables and inferred coefficients ($\hat{\beta}$) is primarily reflective of the portions of the continent that are
better constrained by the 2004 measurement network. For example, under-sampled arid regions in the Western half of the continent may have additional significant flux drivers, and/or somewhat different relationships between driving variables and CO$_2$ flux.

Table 2 presents the biospheric variables that were selected for inclusion in the NARR inversion using the BIC approach described in Sect. 2.5.1 and Supplement E, and their inferred \( \hat{\beta} \) values. Both the selected variables, and their relationship to CO$_2$ flux, are very consistent with process-based understanding of CO$_2$ flux. For example, evapotranspiration explains the predominant portion of the uptake signal (as indicated by a large negative \( \hat{\beta} \)), which is consistent with the strong physiological correlation between plant photosynthetic and transpiration fluxes (Bonan, 2008), and similar relationships between evapotranspiration and CO$_2$ flux have been found in other regression studies using eddy-covariance measurements (e.g. Mueller et al., 2010; Yadav et al., 2010). Combining evapotranspiration estimates with basin-wide estimates of water-use efficiency has also proven to be a robust method for estimating gross primary production (GPP) at watershed scales (Beer et al., 2007). Canopy conductance, which has a similar mechanistic correlation with photosynthesis, was not selected as a significant variable, potentially due to the difficulty in spatially up-scaling this value to landscape scales (Anderson et al., 2003).

The positive \( \hat{\beta} \) values associated with precipitation rate, 30-day average precipitation and specific humidity are consistent with known drivers of heterotrophic respiration (Ise and Moorcroft, 2006). The source associated with precipitation rate at a 3-hourly timescale is consistent with flux tower studies showing pulses of respiration following rain events (Baldocchi, 2008), while the 30-day average precipitation helps to better explain respiration fluxes associated with longer-term soil moisture and water availability. Specific humidity, or the mass of water vapor per unit mass of air, scales with both water availability and temperature, and can therefore additionally help to represent the well-known temperature dependence of respiration (Lloyd and Taylor, 1994). Snow cover is associated with a reduction in respiration sources, consistent with process-based
studies showing that snow can trap soil respiration fluxes from release to the atmosphere until the spring thaw (Kelley et al., 1968; Björkman et al., 2010).

The a posteriori correlation between the $\hat{\beta}$ uncertainties in Table 3 (derived from $V_{\hat{\beta}}$) shows that the influence of the selected variables and their inferred $\hat{\beta}$’s are not completely independent. Therefore, the magnitude of the $\hat{\beta}$’s must be interpreted in combination with other collinear variables. For example, specific humidity corrects the signal associated with evapotranspiration, reducing uptake in warm and humid areas where respiration costs are higher. The $\hat{\beta}$ on percent snow cover is also correlated with that on specific humidity, likely due to an anti-correlation in their underlying values (i.e. cold and snowy areas tend to be less humid).

It should be noted here that not all processes that affect CO$_2$ flux can be easily included in a statistical regression model like this one, especially discrete events that are not reflected in the auxiliary variables available for analysis. For example, this study only chose to examine variables from the NARR, and excluded other possible datasets that could help to explain daily or weekly fluxes (e.g. a fire emission inventory). This choice was made because of the need to represent variability at sub-diurnal scales. Also, while evapotranspiration and canopy conductance implicitly incorporate a measure of live biomass (specifically the Normalized Difference Vegetation Index), there are no variables within the NARR superset that can properly represent substrate availability for respiration, e.g. crop residues after agricultural harvesting or downed woody debris following a storm. In addition, the NARR variables themselves have known limitations (e.g. Bukovsky and Karoly, 2007; West et al., 2007; Markovic et al., 2009). By design, however, any portion of the flux variability that is visible in the atmospheric observations but that cannot be represented using the available auxiliary environmental datasets can still be included in the best estimates in GIM through the spatially-correlated stochastic component of the best estimate (see Eq. A6 in Supplement A). This thereby alleviates the risk posed by excluding certain environmental datasets from the analysis.
3.2 Seasonal cycle of biospheric fluxes

This section presents flux estimates from the Simple and NARR inversions, aggregated from their native 3-hourly resolution to the monthly seasonal cycle, at both grid and spatially aggregated scales. In order to isolate the biospheric portion of the inferred flux, the fossil fuel inventory dataset is subtracted a posteriori from flux estimates from the Simple inversion, whereas the influence of the fossil fuels was already pre-subtracted from the measurements in the NARR inversion (Sect. 2.5.2).

While flux estimates at the grid-scale have high uncertainties, a visual assessment of grid-scale spatial patterns can help to assess the overall results of the inversion, and therefore the strengths and weaknesses of the inversion setup and data constraint for this year. For reference, GIM results are compared to the median of the biospheric models participating in the NACP RCIS (Huntzinger et al., 2011c). Although the biospheric model median masks the large spread in individual biospheric model results, it does represent the current consensus of the terrestrial ecosystem modeling community, and can therefore be used to evaluate the inversion results for obvious discrepancies. In cases where the inversion differs from the biospheric model median, we attempt to interpret what is driving those differences, and which set of estimates may be more reliable. Although the median was chosen here over the mean to reduce the influence of outliers, the across-model mean (using a different set of models) was shown to have relatively high skill in model-data validation studies at the scale of individual flux towers (\(\sim 1 \text{ km}^2\)) as compared to specific model results (e.g. Schwalm et al., 2010). In fact, the monthly spatial patterns from GIM agree more closely with both the biospheric model mean and median than with any one individual model. It is important to note, however, that the biospheric model median may be subject to systematic errors across models, e.g. due to the lack of crop-specific parameterizations or missing processes in agricultural areas, or biases in the driver data.

At spatially aggregated scales, tests conducted with synthetic data and perfect transport showed that the inversion results are mostly robust in well-constrained areas of the
continent. Also, the timing of strong sources and sinks throughout the year is also robust. Therefore, for the spatially aggregated seasonal cycle, we primarily use the GIM results to assess the biospheric models, although the possibility of biases in the GIM estimates cannot of course be entirely eliminated even at these aggregated scales. GIM flux estimates are also compared at this scale to estimates from other inversion studies, in order to assess the impact of inversion inputs and setup (i.e. data, transport, flux resolution, etc.) on flux estimates. If the spread in the inversions across studies were smaller than that seen in the biospheric models, this would provide additional confirmation that the atmospheric data constraint can be used to evaluate bottom-up biospheric model estimates.

3.2.1 Monthly grid-scale spatial patterns

Figure 3 shows the grid-scale fluxes for four months from the Simple and NARR inversions, the biospheric model median for the same months, as well as the Root Mean Squared Difference (RMSD) and grid-scale spatial correlation values between the inversion and biospheric estimates.

Overall, the relatively smooth grid-scale flux estimates from the Simple inversion represent the diffuse information content of the available atmospheric CO$_2$ data in 2004. The smoothness is indicative of both the sparseness of the network, as well as atmospheric mixing that dilutes the influence of surface sources and sinks. In addition, the isotropic spatial covariance assumptions in $Q$ spread out the flux signal equally in all directions, whereas correlations are most likely shorter in the latitudinal direction relative to the longitudinal.

The NARR inversion recovers more realistic grid-scale heterogeneity in the flux estimates, as can be seen by stronger correspondence with the biospheric model median. The NARR inversion may also reduce the impact of the isotropic $Q$, given that the NARR variables can help to explain many of the large-scale features in the flux distribution, thereby accounting for the stronger longitudinal correlations in flux. Some features of the NARR inversion results, such as the strong sinks in the Southeastern
continental United States and the coastal Pacific forests during the growing season, are likely due to the extrapolation of process-based relationships from better constrained areas. In some cases, this extrapolation may not be realistic, such as the sources in the NARR inversion in most months in the Northwest coastal areas, which reflect spatial patterns from the evapotranspiration variable, but that are not consistent with the biospheric model median.

One can see that both inversions, but especially the NARR inversion, have a closer correspondence with the biospheric model median during the growing season relative to the dormant season. For example, the correlation coefficient between the NARR inversion estimates and the biospheric model median is 0.72 and 0.65 respectively in April and July, compared to only 0.38 and −0.10 in January and October, respectively. While this increased correspondence during the growing season is not conclusive proof that either the inversions or the biospheric models are correct, it seems likely that the biospheric fluxes in the spring and summer have a stronger signal that is easier to identify from the atmospheric measurements, and also that there is a greater availability of auxiliary variables that can help to explain fluxes during this time of the year. For example, evapotranspiration, associated with uptake, explains the predominant portion of the variability in the NARR inversion. Conversely, the variables helping to explain respiration sources, which dominate NEE in the winter months, have a weaker contribution to the final estimates. The biospheric models themselves may also have increased skill during the growing, relative to the dormant, season, as seen in other model intercomparison studies using eddy-covariance data (e.g. Schwalm et al., 2010).

During the dormant season, there are a number of potential factors that can help to explain the discrepancies between flux estimates from the inversions and the biospheric model median: (1) faster wind speeds during the winter months (Elliott et al., 1986; CWEA, 2011), particularly in January shown here, may be leading to more weakly-defined spatial patterns from the inversions, which also complicates their correlation with NARR variables. (2) Errors in the magnitude or spatial patterns of the fossil fuel inventories would have more impact on the inversion results in the dormant season.
when these emissions form a larger proportion of the total CO$_2$ flux. (3) Most of the biospheric models in the NACP RCIS do not specifically account for managed agricultural processes (i.e. fertilizer, irrigation, crop harvest and transport, etc.), which likely reduces model skill in agricultural areas of the continent (Lokupitiya et al., 2009; Corbin et al., 2010; Huntzinger et al., 2011c). This is particularly evident in October when the strong sources in the center of the continent from the inversions are inconsistent with the biospheric model median. These inversion sources reflect a strong build-up of CO$_2$ relative to background air at most towers in the continental United States, particularly at the Park Falls, Wisconsin and Norman, Oklahoma measurement locations, and are most likely associated with the decay of residual biomass after crop harvesting (Johnson et al., 2006). Similarly, one can see stronger sinks in the agricultural Midwest during July in the inversions relative to the biospheric model median. The inversion sinks are supported by process-based studies showing stronger productivity in well-irrigated and fertilized agricultural croplands relative to unmanaged grasslands, or the native vegetation of these areas as parameterized in many of the biospheric models (Lokupitiya et al., 2009; Corbin et al., 2010; Smith et al., 2010).

### 3.2.2 Biome-scale seasonal cycle

Figure 4 presents fluxes aggregated both temporally, to monthly means, and spatially, to the three biomes and continent shown in Fig. 2b. The Simple and NARR geostatistical inversions are first compared to one another, and then to previous inversion studies (left column of Fig. 4), as well as to individual biospheric models (right column of Fig. 4). The biospheric model with the closest agreement with the GIM estimates is also highlighted for each region.

Inclusion of NARR auxiliary variables is seen to have little impact here on flux estimates at aggregated spatial scales, particularly in areas with a strong atmospheric data constraint, consistent with results from Gourdji et al. (2008). For example, aggregated flux estimates from the two GIM inversions are almost identical in the Boreal Forest, Temperate Grass/Savannah/Shrub, and for the full continent, despite significant
differences in small-scale spatial patterns (Fig. 3). In the Eastern Temperate Forest, which has a weaker atmospheric constraint, the NARR auxiliary variables have more of an impact on estimates, with an earlier start to the growing season and a stronger peak uptake relative to the Simple inversion. This stronger uptake could be realistic given that the under-constrained Southeastern United States forests are known to be very productive (Baker et al., 2010; Crevoisier et al., 2010). Overall, these results show that in areas with a strong atmospheric data constraint, the NARR variables are unlikely to bias estimates at aggregated spatial scales, while hopefully improving them at finer scales and in under-constrained regions.

The comparison to other inversion studies shows that the GIM recovers a seasonal cycle that is within the spread of the other estimates. The spread across inversions is narrower in the better-constrained biomes (i.e. Temperate Grass/Savannah/Shrub and Boreal Forest), and wider in the Eastern Temperate Forest. This result points to the value of an expanded measurement network for reducing the impact of inversion setup choices and assumptions on final flux estimates. The spread between inversions is also narrowest for the two largest examined regions (Boreal Forest and the full continent), indicating a better constraint on flux estimates from inversions at progressively larger spatial scales, at least with the limited 2004 measurement network.

Flux estimates from GIM for the spatially-aggregated seasonal cycle are similar to those from CarbonTracker, particularly in the Eastern Temperate and Boreal Forests, despite significantly different flux resolutions and transport models (Table 1), while the Schuh et al. (2010) and Butler et al. (2010) estimates diverge more strongly from the other results. The strong growing season uptake in the Butler et al. (2010) study relative to the other inversions may be due to spatial aggregation errors associated with using measurements from highly productive areas, and extrapolating this signal too strongly and evenly across the coarse estimation regions (see Table 1). Also, the monthly concentration averaging intervals and flux estimation timescale used in the Butler et al. (2010) study may also lead to exaggerated sink strengths relative to inversions using finer temporal resolutions. Both spatial and temporal aggregation errors could
be affecting the CarbonTracker results as well in the Temperate Grass/Savannah/Shrub biome, where this inversion shows stronger peak uptake in July and August relative to the other studies.

The Schuh et al. (2010) results show a dampened seasonal cycle and an earlier peak uptake relative to the other inversions in the Eastern Temperate Forests and Temperate Grass/Savannah/Shrub. This result could be due to a strong adherence to the seasonal cycle of the prior (SiB3), which also has relatively early peak uptake in the continental US relative to other biospheric models, e.g. CASA-GFEDv2, and/or other unique aspects of the setup for this study (see Table 1 and original publication).

It should be noted that the Butler et al. (2010) flux estimates using CASA vs. SiB3 as a prior differ only marginally at this aggregated scale, similarly to the geostatistical inversion results using different flux covariates. This suggests that prior flux assumptions may not be driving the observed spread across inversion studies at aggregated scales in well-constrained areas of the continent as much as other aspects of the inversion setups (e.g. flux spatial and temporal resolution, concentration averaging timescales, atmospheric transport models, and flux covariance assumptions).

The spread in the biospheric models for the aggregated seasonal cycle is slightly larger than that seen across inversion studies, especially in the Temperate Grass/Savannah/Shrub biome (i.e. the agricultural areas of the continent), and in the forested biomes during the growing season, where the magnitude of peak uptake varies strongly among the models. The GIM results mostly fall within this spread, except in the Temperate Grass/Savannah/Shrub biome, where the inversions show a stronger peak uptake shifted a month later compared to the majority of the biospheric models. GIM also shows strong sources to the atmosphere in this biome in March and October, which are lacking in most of the biospheric models, although the other inversion studies have weaker sources in these months relative to GIM. The lack of agreement in this biome between GIM and the biospheric models may again reflect the fact that the biospheric models participating in the NACP RCIS generally do not include crop-specific parameterizations and may omit relevant processes, e.g. due to land-management activities,
which are required to realistically simulate NEE in agricultural areas (Lokupitiya et al., 2009; Corbin et al., 2010).

In the Boreal Forests and at the continental scale, the spread in the biospheric models is narrower than in the other two biomes and more consistent with the GIM results. In the Boreal Forests, despite a difference in the magnitude of flux estimates, all models agree reasonably well on the timing of the seasonal cycle. At the continental scale, the spread in both the biospheric model estimates and the inversion studies is the narrowest, pointing to errors that cancel at large scales in both bottom-up and top-down models for the seasonal cycle.

Comparing biome-scale GIM estimates to those from individual biospheric models, in addition to the full ensemble of models, can potentially help to highlight strengths and weaknesses in different bottom-up modeling approaches. The GIM estimates are strongly driven by the information content of the atmospheric data, which provide a relatively strong constraint on fluxes at aggregated spatial scales. Therefore, comparing the GIM estimates to individual biospheric models can highlight the impact of these models’ parameterizations on large-scale fluxes. For example, in the Eastern Temperate Forest biome, CASA-GFEDv2 shows remarkably close agreement with the Simple inversion from January to September, although it has weaker uptake than the NARR inversion from April to August. If the NARR inversion were indeed more realistic in this productive area (especially the under-sampled Southeastern forest plantations), this could imply that CASA-GFEDv2 shows a late start to the growing season, as discussed in Randerson et al. (2009). The two biospheric models with the strongest growing season uptake in this biome, EC-MOD and MOD17+, both diagnostic models, appear to be unrealistic when compared to GIM estimates.

In the Temperate Grass/Savannah/Shrub biome, EC-MOD shows strong correspondence with GIM from April through August in terms of both the timing and magnitude of uptake. EC-MOD, somewhat similarly to the NARR inversion, extrapolates spatial patterns of flux with auxiliary environmental datasets and derived statistical relationships, although the extrapolation is done using eddy-covariance flux tower measurements.
rather than in an atmospheric inversion framework (Xiao et al., 2008). The empirical nature of EC-MOD reduces the impact of misrepresented processes in agricultural areas, which likely bias estimates in the other biospheric models. However, EC-MOD shows weaker sources than GIM throughout the dormant season, particularly in March and October. Only DLEM shows similarly strong sources in October, although with the peak source shifted one month later to November. It is difficult to know if DLEM is matching the inversion in the autumn months for the correct reasons, but this result points to the need for the other biospheric models to more carefully represent decomposition of agricultural residue following the harvest.

Overall, this comparison across inversions and biospheric models for the aggregated biome-scale seasonal cycle indicates that: (1) prior flux assumptions may not be the strongest driver of the spread in the inversion results at aggregated spatial scales, as seen by the similarity between the Simple and NARR geostatistical inversion flux estimates, and consistent with conclusions from Butler et al. (2010), (2) the spread across inversion studies is somewhat reduced in better-constrained areas and at larger spatial scales, although aggregation errors, particularly in the Butler et al. (2010) results, appear to be driving the largest portion of this spread, (3) the great majority of the biospheric models included here appear to lack skill in the agricultural areas of the continent, and (4) inversions taking advantage of continuous, continental datastreams, particularly GIM shown here, hold promise for evaluating biospheric model output at the scale of sub-continental biomes in future work.

### 3.3 Net annual sources and sinks

Net annual sources and sinks represent the small residual on a strongly-varying seasonal cycle; therefore, for all modeling approaches, uncertainties which may seem reasonable at the monthly timescale will represent a larger proportion of the annual totals. Also, biases in the inversions or biospheric models at short timescales, e.g. associated with driver data, may accumulate at longer timescales. For these reasons, the annual flux estimates presented here, both from GIM and the other model-based estimates,
are considered to be a less robust result than those for the monthly seasonal cycle, although we still expect the inversions to be fairly reliable at the annual timescale in well-constrained areas, based on synthetic data experiments. Also, net annual fluxes from GIM are presented with both the CarbonTracker and GlobalView boundary conditions, which were seen to have a cumulative impact on flux estimates across the year due to their consistent offset. (See Supplement B for a discussion of the offset between the boundary conditions.)

3.3.1 Annual grid-scale spatial patterns

Figure 5 shows the annual grid-scale sources and sinks from the Simple and NARR inversions with both sets of boundary conditions, as compared to the biospheric model median. The boundary conditions have a strong impact on the magnitude of annual flux estimates. This can be seen most clearly in the Simple inversion results, where the results using CarbonTracker boundary conditions show much stronger uptake in the boreal areas and weaker sources in the desert Southwest as compared to flux estimates with GlobalView boundary conditions. An analysis of monthly grid-scale flux estimates showed that most of these differences at the annual scale are due to a difference in net uptake from March through August. Differences were smaller but still evident outside of the growing season.

The inclusion of NARR variables adds significant detail to the annual spatial patterns from GIM, particularly in under-constrained areas, and also increases the strength of both sources and sinks. For example, the NARR inversions show substantially higher uptake in the Eastern US as compared to both the Simple inversion results and the biospheric model median. The locations of net uptake from the NARR inversions, in the Eastern US and the Midwestern agricultural areas across the US and Canada, are in relatively close agreement with bottom-up inventory estimates from the State of the Carbon Cycle Report (SOCCR; CCSP, 2007), as well as top-down estimates of Crevoisier et al. (2010), which, while not an inversion study, used an independent
carbon budgeting method for North America based on vertical profiles of CO₂ mixing ratios collected from aircraft samples over the continent.

In the under-constrained areas, inversion results are likely to be less reliable. For example, the strong net sources in the NARR inversions in the desert Southwest are mostly inconsistent with the biospheric model median, and may reflect a dipole from the uptake in the Eastern continental US. However, the sign of the flux in these areas may still be accurate. For example, individual biospheric models like EC-MOD do show sources in these areas with a reduced magnitude. Also, Hayes et al. (2011) found weak sources in these regions using an inventory-based approach. Given sparse vegetation, Hayes et al. (2011) and Crevoisier et al. (2010) hypothesized that livestock consumption of agricultural products grown elsewhere could be contributing to sources in these arid pasture-lands. The lack of these sources in the majority of the biospheric models is consistent with the fact that many of these do not explicitly account for the lateral transport of agricultural products.

### 3.3.2 Aggregated annual budgets

Figure 6 shows annual flux estimates spatially aggregated to the same regions as in Fig. 4, for GIM as well as the other inversions and the full suite of available biospheric models, although biospheric models and inversions with less than 85% coverage for a given region were excluded from the comparison (e.g. Schuh et al., 2010 at the continental scale). Contrary to the other figures, Fig. 6 shows the total CO₂ flux including the fossil fuel component, with a line indicating the estimated fossil fuel emissions from the combined emission inventory dataset used in this study. While fossil fuel emissions are generally considered to be well-known, especially at the annual scale, there still remains some uncertainty associated with the magnitude of these estimates (e.g. Francey et al., 2010). Therefore, we show total CO₂ flux, as this is the signal actually seen by the atmospheric measurements, and is therefore the most robust result from the inversion.
In the Eastern Temperate Forests at the annual aggregated scale, all inversions and biospheric models show net uptake relative to the fossil fuel emission signal. The inversion results for both GIM and other methods show a wider spread in this biome relative to the others, reflecting the partial atmospheric data constraint in this region. In particular, the NARR variables are seen to have a large impact on the GIM results, increasing the net sink relative to the Simple inversion by approximately 0.3 PgC yr\(^{-1}\). In fact, the Simple inversion results for this biome are more consistent with CarbonTracker, Schuh et al. (2010) and the great bulk of the biospheric models, as compared to the NARR inversion. The strong uptake of approximately 1 PgC yr\(^{-1}\) seen in Butler et al. (2010) is inconsistent with all but two of the biospheric models, and may reflect the impact of aggregation errors, as suggested previously.

In the Boreal Forests, the inversion spread in the annual aggregated fluxes (about 0.4 PgC yr\(^{-1}\)) is smaller than that in the Eastern Temperate Forests (about 0.9 PgC yr\(^{-1}\)), and all studies show net uptake. Also, the flux estimates from the Simple and NARR inversions are relatively similar, with only a slightly weaker sink from the NARR inversion. While there is some consistency between the inversions and the biospheric models, particularly for GIM using GlobalView boundary conditions, the spread in the biospheric model output is shifted more towards a neutral flux relative to that seen in the inversions for this biome.

The Temperate Grass/Savannah/Shrub is the biome best-constrained by the atmospheric data, and not surprisingly, the differences between the GIM estimates is smallest in this region, with the Simple and NARR inversions returning an almost identical flux. The other inversion studies are also mostly consistent with one another, but show a stronger uptake relative to GIM. It is difficult to say which set of inversions is more “correct” in this biome, but it is likely that the stronger sources in the dormant season in GIM relative to the other studies, particularly in March and October (Fig. 4, Sect. 3.2.2), can explain this difference in net annual flux across studies. Given that the biospheric models for the most part lacked these same sources, it could be that the other inversion studies are too closely tied to their bottom-up priors in these agricultural regions, and
not sufficiently sensitive to the signal in the atmospheric data. It may also be that the sources in the dormant season in GIM are too strong, for currently unknown reasons.

For the North American continental budget, the inversion and biospheric model spread is quite large compared to the budget at sub-continental scales. The annual continental spread also appears to be larger than that for the spatially-aggregated continental seasonal cycle (Fig. 4), although this partially reflects the different unit for the annual aggregated budgets in Fig. 6 (PgC yr\(^{-1}\)) relative to that used for the seasonal cycle in Fig. 4 (µmol m\(^{-2}\) s\(^{-1}\)).

At least for GIM, the largest contributor to the spread in the annual continental budget is the influence of the two sets of boundary conditions, which appears to have an additive effect in both space and time. In fact, using the GlobalView boundary conditions, GIM returns an almost neutral biospheric flux, and reduces the net North American sink by approximately 0.7 to 0.9 PgC yr\(^{-1}\) relative to the GIM inversion using CarbonTracker boundary conditions. Interestingly, CarbonTracker (Peters et al., 2007) itself finds a net sink almost identical to that seen in the NARR inversion with CarbonTracker boundary conditions. This may suggest that the CO\(_2\) concentrations of air flowing into the continent are the primary determinant of the apparent continental budget from top-down inversion studies, rather than the flux resolution, priors, transport model or data choices. An inventory-based estimate of the North American carbon balance from Hayes et al. (2011) finds a net biospheric uptake of 0.28 PgC yr\(^{-1}\), which is more similar to the GIM results using the GlobalView boundary conditions. However, the main conclusion to take away here is that relatively low uncertainties on these boundary conditions are required in future work to accurately pinpoint the net biospheric uptake from top-down inversion studies based on the atmospheric constraint.

4 Conclusions

A geostatistical inversion approach was implemented here to quantify North American fluxes of CO\(_2\) for 2004 directly at a 1° × 1° spatial, and 3-hourly temporal, resolution.
Covariance parameters and process-based information included in the inversion were optimized using the atmospheric data constraint. By estimating fluxes at finer scales relative to previous inversions, this study reduces spatial and temporal aggregation errors associated with using continental measurements sited in areas with high flux variability. In addition, avoiding the use of explicit prior flux estimates allows for a more independent comparison to flux estimates from biospheric models. Finally, the inferred regression coefficients associated with climatological datasets included in the inversion provide useful inference regarding the relationship between significant regional flux drivers and inferred flux magnitudes.

Results show that the estimated drift coefficient on the fossil fuel inventory in the Simple inversion was near one, providing support for the quality of the inventory as well as the GIM setup and assumptions. Also, the introduction of NARR auxiliary variables into the inversion yielded regression coefficients that were very consistent with process-based understanding of the drivers of component CO\textsubscript{2} fluxes from photosynthesis and respiration, with evapotranspiration explaining a substantial portion of the net uptake signal. These variables were additionally seen to reduce the impact of assuming an unrealistic isotropic spatial covariance structure on grid-scale flux estimates, and to help map fluxes in underconstrained regions by extrapolating the flux signal from the atmospheric data. In addition, at aggregated spatial scales in well-constrained regions of the continent, the Simple and NARR inversions produced similar flux estimates, pointing to the atmospheric data as the primary constraint on fluxes at larger scales.

The top-down/bottom-up inter-comparison implemented in this study provided a new approach for evaluating biospheric CO\textsubscript{2} flux estimates at regional scales, by providing insights into the strengths and weaknesses of both the examined inversion studies and biospheric models. For example, the inversions were found to have more consistent spatial patterns with biospheric models during the growing season relative to the dormant season. This could be due to higher wind speeds in the winter and a more diffuse flux signal, errors in the fossil fuel inventories that are aliased onto the inferred fluxes, which would be more evident in the dormant season when emissions dominate the
total CO$_2$ flux, or errors in the biospheric models themselves which may have less skill outside of the growing season (e.g. Schwalm et al., 2010). Finally, the fluxes during the growing season are stronger and more variable, and therefore perhaps easier to identify from the atmospheric signal.

At the biome-scale where inversion flux estimates are presumably more robust than at the grid-scale, comparisons of inferred seasonal cycles pointed to the strong impact of inversion setup on flux estimates, particularly aggregation errors associated with estimating fluxes at coarse-scales relative to the fine-scale variability embedded in bottom-up prior flux estimates. A strong atmospheric data constraint reduced this impact somewhat, with inversions that estimate fluxes at different spatial scales showing more consistent results in well-constrained regions, i.e. the Boreal Forests and Temperate Grass/Savannah/Shrub, pointing to the value of an expanded measurement network for reducing the sensitivity of inversion results to setup choices. Prior flux assumptions were seen to be less important in contributing to the inversion spread at this scale than other inversion setup choices (flux resolution, concentration averaging intervals, atmospheric transport, etc.) Controlled experiments varying inversion components individually would make it possible to more conclusively highlight the impact of inversion setup assumptions and input choices.

The comparison of GIM results to the biospheric models at the biome-scale pointed to the need for the biospheric models to better account for land management activities, crop parameterizations and the fate of harvested products in the agricultural areas of the continent. For example, the geostatistical inversions showed strong sources in the center of the continent in March and October. While the magnitude of these sources was somewhat sensitive to inversion setup, a peak in respiration sources in the same regions during these months was not seen in the majority of the biospheric models. The fact that these sources were also absent in some of the examined inversion studies (CarbonTracker and Schuh et al., 2010) may point to the impact of these biospheric model limitations on a posteriori flux estimates for inversions based on these models. The comparison of the GIM results to individual biospheric models was not definitive as
to the performance of either type of model, but provided some hope for using regional inversion output to validate biospheric models in future work.

Finally, accurate annual carbon budgets at regional scales (e.g. $1^\circ \times 1^\circ$) from either regional inversions or biospheric models appear to remain an elusive goal at the current time. While there appears to be some convergence in the location of net uptake and release over the continent from the geostatistical inversions, the biospheric model median, the SOCCR report (CCSP, 2007), and Crevoisier et al. (2010), grid-scale flux estimates at the annual scale were very different between the Simple and NARR inversions, and also as compared to the biospheric model median.

At aggregated spatial scales, annual fluxes from the Simple and NARR inversions were more consistent, particularly in well-constrained biomes. However, in the Boreal Forests and Temperate Grass/Savannah/Shrub, the inversions and biospheric models did not even agree on the sign of the flux. Moreover, at the continental scale, the boundary conditions used as input into regional inversions were found to lead to large differences in estimated continental flux.

Overall, the North American geostatistical inversion for 2004 presented here provides a robust inversion framework suitable for ingesting the large data volumes associated with the recent expansion of the in situ CO$_2$ monitoring network over North America (Mueller et al., 2011). In particular, the presented GIM approach estimates fluxes at unprecedented spatiotemporal scales that help to optimally take advantage of the information contained in highly variable continental measurement data. Top-down/bottom-up comparison studies like this one can also provide insight into the currently large spread of biospheric and inverse model estimates of regional CO$_2$ flux, thereby providing a path forward for improved estimates in future work.

Supplementary material related to this article is available online at: http://www.biogeosciences-discuss.net/8/6775/2011/bgd-8-6775-2011-supplement.pdf.
Acknowledgements. This work was supported by NASA ROSES Grant # NNX06AE84G, “Constraining North American Fluxes of Carbon Dioxide and Inferring Their Spatiotemporal Covariances through Assimilation of Remote Sensing and Atmospheric Data in a Geostatistical Framework”, and a NASA Earth System Science Fellowship (awarded to S. Gourdji).

We would also like to thank biospheric modelers who contributed results to the NACP Regional and Continental Interim Synthesis and Mac Post and Yaxing Wei of Oak Ridge National Laboratory who helped to organize and process this model output. We also thank the other inverse modelers (Andrew Schuh, Martha Butler, Wouter Peters, Andy Jacobson and co-authors) who contributed model output and valuable interpretations for the inter-comparison. CarbonTracker 2009 results were provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov.

Other data providers (not included as co-authors) include Doug Worthy of Environment Canada for the Canadian continuous measurement sites, and Bill Munger from Harvard University for the Harvard Forest CO$_2$ data.

Finally, The WRF-STILT development at AER has been funded by the National Science Foundation Atmospheric Chemistry Program (grant # ATM-0836153). Steve Wofsy and Christoph Gerbig are thanked for their continuing contributions to the STILT model development. The WRF and STILT runs described in this paper have been made possible by access to NASA’s high-end computing resources and we thank the personnel at the NASA Ames supercomputing facility for technical assistance.

References


Crevoisier, C., Sweeney, C., Gloor, M., Sarmiento, J. L., and Tans, P. P.: Regional US carbon


Gurney, K., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L.,


International Land Model Benchmarking Project – ILAMB, http://www.ilamb.org/, last access: 2

6816
June 2011.


Oda, T. and Maksyutov, S.: A very high-resolution (1 km $\times$ 1 km) global fossil fuel CO$_2$ emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543–556, doi:10.5194/acp-11-543-2011, 2011.


Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer,
Schuh, A. E., Denning, A. S., Uliasz, M., and Corbin, K. D.: Seeing the forest through the trees: Recovering large scale carbon flux biases in the midst of small-scale variability, J. Geophys.


Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn,
Table 1. Comparison of setup and input data across published inversion studies for North America in 2004.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Spatial and temporal resolution</th>
<th>A priori flux covariance</th>
<th>Prior – biospheric</th>
<th>Prior – fossil fuels</th>
<th>Winds and transport model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butler et al. (2010)</td>
<td>Global 10 sub-regions in North America, monthly</td>
<td>None</td>
<td>SiB3 (hourly)/CASA (monthly mean)</td>
<td>Erickson et al. (2008)</td>
<td>GEOS-4/PCTM</td>
</tr>
<tr>
<td>Schuh et al. (2010)</td>
<td>North America 1° x 1°, weekly</td>
<td>Spatial covariance using fixed correlation length scales</td>
<td>SiB3 (hourly)</td>
<td>Vulcan in continental US (Gurney et al., 2009); Andres et al. (1996) elsewhere</td>
<td>RAMS/LPDM</td>
</tr>
<tr>
<td>GIM – Simple</td>
<td>North America 1° x 1°, 3-hourly</td>
<td>Monthly-varying spatial covariance using correlation length scales and variances estimated with atmospheric data</td>
<td>None</td>
<td>Vulcan in continental US (Gurney et al., 2009); CDIAC/Night Lights dataset elsewhere (Oda and Maksyutov, 2011)</td>
<td>WRF/STILT</td>
</tr>
<tr>
<td>GIM – NARR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6824
Table 2. Candidate NARR environmental covariates, with associated $\hat{\beta}$ values for those variables selected using BIC for inclusion in the NARR inversion with empirical boundary conditions. Auxiliary variables were normalized to zero mean and unit variance, such that $\hat{\beta}$ values are directly comparable. Variables with dashes (“−”) were considered but not selected by the BIC/Branch and Bound algorithm. Also shown are the coefficients of variation (CV) for the selected variables (i.e. $|\sigma/\hat{\beta}|$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\beta}$</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canopy conductance</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Downward shortwave radiation</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>−1.54</td>
<td>0.04</td>
</tr>
<tr>
<td>Precipitation rate</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Specific humidity</td>
<td>0.10</td>
<td>0.37</td>
</tr>
<tr>
<td>Soil moisture</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Air temperature (@ 2 m)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Plant canopy water content</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Snow depth</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Snow cover (%)</td>
<td>−0.13</td>
<td>0.18</td>
</tr>
<tr>
<td>16-day average precipitation</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>30-day average precipitation</td>
<td>0.18</td>
<td>0.19</td>
</tr>
</tbody>
</table>
**Table 3.** Correlation coefficients among $\hat{\beta}$ uncertainties for variables included in the NARR inversion with empirical boundary conditions.

<table>
<thead>
<tr>
<th></th>
<th>Evapotranspiration</th>
<th>Precipitation rate</th>
<th>Specific humidity</th>
<th>Snow cover (%)</th>
<th>30-day average precipitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evapotranspiration</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Precipitation rate</td>
<td>0.17</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Specific humidity</td>
<td>–0.44</td>
<td>–0.29</td>
<td>1.00</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Snow cover (%)</td>
<td>0.11</td>
<td>–0.06</td>
<td>0.46</td>
<td>1.00</td>
<td>–</td>
</tr>
<tr>
<td>30-day average precipitation</td>
<td>–0.22</td>
<td>–0.17</td>
<td>–0.21</td>
<td>–0.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Fig. 1. Domains of nested WRF winds, flux estimation grid, and the locations of towers, flask and aircraft measurements used in the inversions. Please see Table B1 of the Supplement for a key to the tower names.
Fig. 2. (a) Yearly-averaged sensitivity to fluxes for all observation locations. (b) Three biomes (modified from Olson, 2001) used for spatial aggregation of flux estimates. These three biomes, along with the grey grid-cells, form the domain for the aggregated North American totals in Figs. 4 and 6.
Fig. 3. Monthly-averaged grid-scale biospheric fluxes from the Simple and NARR inversions using GlobalView boundary conditions, and the median of biospheric models participating in the NACP RCIS. Also shown are the grid-scale spatial correlation coefficients ($\rho$) and Root Mean Squared Difference (RMSD; $\mu$mol m$^{-2}$ s$^{-1}$) between the inversion and the biospheric model median fluxes for each month. Please note the different scales per month.
Fig. 4. Seasonal cycle of monthly-averaged fluxes aggregated to the three relatively well-constrained biomes, as well as the full continent (Fig. 2b). GIM fluxes using empirical boundary conditions are compared to other inversions and biospheric models that have coverage in at least 85% of the given regions. Inversions with the same line color use similar biospheric model output for their prior flux estimates. All biospheric models are shown in the right panel in grey lines. The biospheric model having the closest agreement with the GIM fluxes at this scale is highlighted for each region.
Fig. 5. Annually-averaged grid-scale biospheric fluxes from the Simple and NARR inversions using GlobalView and CarbonTracker boundary conditions, and the median of biospheric models participating in the NACP RCIS. Also shown are the grid-scale spatial correlation coefficients ($\rho$) and Root Mean Squared Difference (RMSD; $\mu$mol m$^{-2}$ s$^{-1}$) between the inversion and the biospheric model median fluxes.
Fig. 6. Annual total flux estimates spatially aggregated to the three biomes and North America (Fig. 2b). Results from the Simple and NARR inversions using empirical and CarbonTracker boundary conditions are compared to results from other inversions and the biospheric models participating in the NACP RCIS. The dotted line indicates the CO$_2$ flux from North American fossil fuel emissions in 2004, as estimated with the combined inventory dataset used in this study (Gurney et al., 2009; Oda and Maksyutov, 2011); i.e. this line corresponds to a neutral biospheric flux.