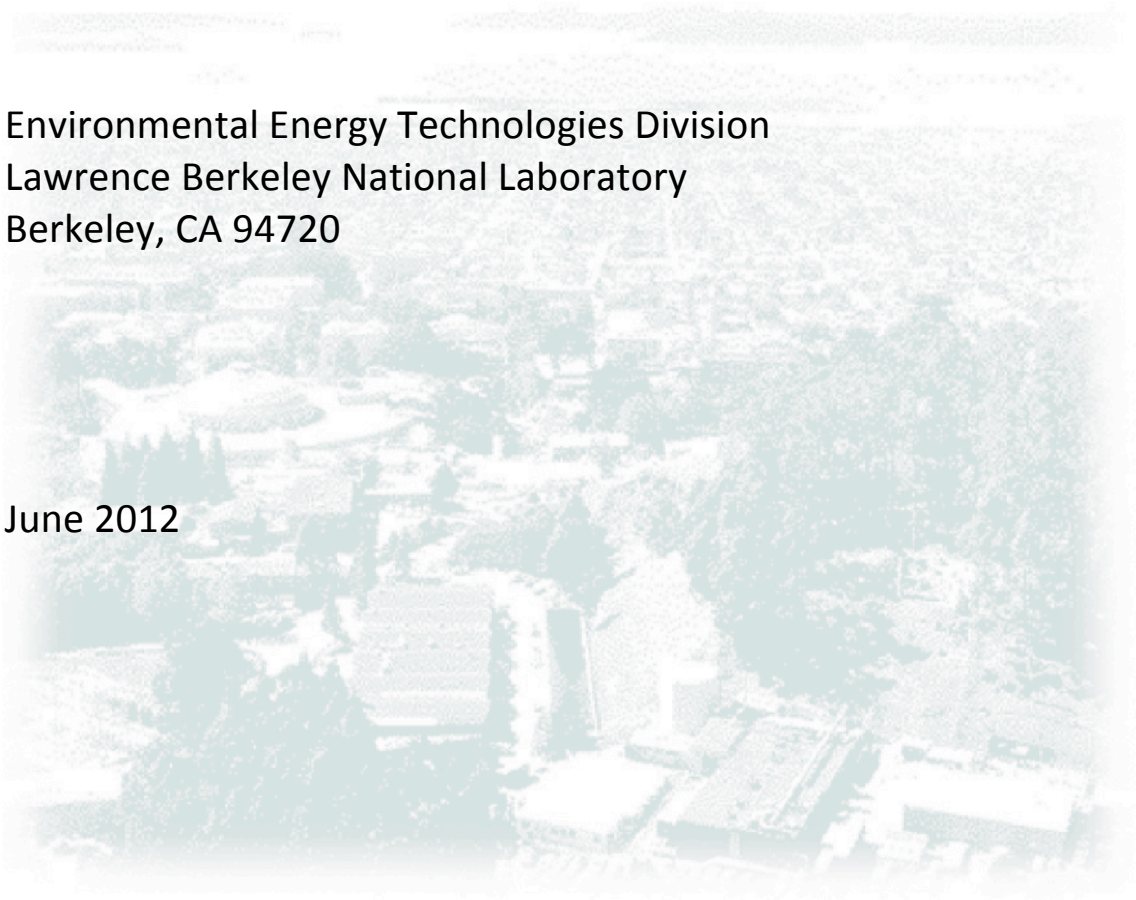
ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

North American CO_2 exchange: Inter-comparison of modeled estimates with results from a fine-scale Atmospheric inversion

Sharon M. Gourdji, Kimberly L. Mueller, Vineet Yadav, Deborah N. Huntzinger, Arlyn E. Andrews, Michael Trudeau, Gabrielle Petron, Thomas Nehrkorn, Janusz Eluszkiewicz, John Henderson, Deyong Wen, John Lin, Marc Fischer, Colm Sweeney, Anna M. Michalak

Environmental Energy Technologies Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94720



June 2012

North American CO_2 exchange: Inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion

Sharon M. Gourdji[1,2], Kimberly L. Mueller[1,3], Vineet Yadav[1,4], Deborah N. Huntzinger[1,5], Arlyn E. Andrews[6], Michael Trudeau[6], Gabrielle Petron[6], Thomas Nehrkorn[7], Janusz Eluszkiewicz[7], John Henderson[7], Deyong Wen[8], John Lin[8], Marc Fischer[9], Colm Sweeney[6] & Anna M. Michalak[1,4]

[1] Department of Civil & Environmental Engineering, University of Michigan, Ann Arbor, MI, 48108, USA
[2] Now at Department of Environmental Earth System Science, and Center on Food Security and the Environment, Stanford University, Stanford, CA 94305, USA
[3] Now an American Association for the Advancement of Science Policy Fellow, Washington, D.C., 20515, USA
[4] Now at Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA
[5] Now at School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ, 86011, USA
[6] Global Monitoring Division, Earth System Research Laboratory, National Oceanic & Atmospheric Administration, Boulder, CO 80305, USA
[7] Atmospheric and Environmental Research, Inc., Lexington, MA, 02421, USA
[8] Department of Earth & Environmental Sciences, University of Waterloo, Waterloo, ON, Canada N2L 3G1
[9] Lawrence Berkeley National Lab, Berkeley, CA, 94720, USA

Corresponding author: Sharon Gourdji, sgourdji@stanford.edu

Abstract

Atmospheric inversion models have the potential to quantify CO_2 fluxes at regional, sub-continental scales by taking advantage of near-surface CO_2 mixing ratio observations collected in areas with high flux variability. This study presents results from a series of regional geostatistical inverse models (GIM) over North America for 2004, and uses them as the basis for an inter-comparison to other inversion studies and estimates from biospheric models collected through the North American Carbon Program Regional and Continental Interim Synthesis. Because the GIM approach does not require explicit prior flux estimates and resolves fluxes at fine spatiotemporal scales (i.e. 1˚x1˚, 3-hourly in this study), it avoids temporal and spatial aggregation errors and allows for the recovery of realistic spatial patterns from the atmospheric data relative to previous inversion studies. Results from a GIM inversion using only available atmospheric observations and a fine-scale fossil fuel inventory were used to confirm the quality of the inventory and inversion setup. An inversion additionally including auxiliary variables from the North American Regional Reanalysis found inferred relationships with flux consistent with physiological understanding of the biospheric carbon cycle. Comparison of GIM results

with bottom-up biospheric models showed stronger agreement during the growing relative to the dormant season, in part because most of the biospheric models do not fully represent agricultural land-management practices and the fate of both residual biomass and harvested products. Comparison to earlier inversion studies pointed to aggregation errors as a likely source of bias in previous sub-continental scale flux estimates, particularly for inversions that adjust fluxes at the coarsest scales and use atmospheric observations averaged over long periods. Finally, whereas the continental CO_2 boundary conditions used in the GIM inversions have a minor impact on spatial patterns, they have a substantial impact on the continental carbon budget, with a difference of 0.8 PgC/yr in the total continental flux resulting from the use of two plausible sets of boundary CO_2 mixing ratios. Overall, this inter-comparison study helps to assess the state of the science in estimating regional-scale CO_2 fluxes, while pointing towards the path forward for improvements in future top-down and bottom-up modeling efforts.

1. Introduction

Carbon cycle scientists are increasingly called upon to provide information in support of efforts to monitor anthropogenic 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 towards these goals by taking advantage of the information contained in atmospheric CO_2 mixing ratio measurements regarding upwind surface 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 CO_2 fluxes. In addition, inverse-modeling derived CO_2 flux estimates are potentially useful for evaluating the process-based formulations of terrestrial ecosystem models. In fact, atmospheric measurements of 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.

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 sub-continental scale fluxes, thereby informing carbon management efforts and evaluations of mechanistic models of the carbon cycle. An expanding *in situ* continuous measurement network across the North American and European continents (e.g. NOAA-ESRL, 2011; CGGMN, 2011; CEAD, 2011) is

making this possible, but optimally extracting the flux signal from these data is complicated by 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.

By limiting the domain size in regional inversions, fluxes can be estimated at relatively fine spatial scales (e.g. 1°x1°), 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 better resolve atmospheric dynamics in the near-field of measurement locations (e.g. Lin et al., 2003; Gerbig et al., 2008). However, it has proven difficult to reconcile observed differences in estimates across regional inverse modeling studies due to a number of potential factors: atmospheric transport, flux spatial and temporal resolution, boundary conditions, errors in the bottom-up models used as priors, treatment of fossil fuels, spatiotemporal flux covariance assumptions, and the processing and filtering of observations, among other possible causes.

Similarly, large discrepancies have also been seen in inter-comparisons of mechanistic models of biospheric CO_2 flux (Huntzinger et al., in review; Hayes et al., in press). Given that atmospheric CO_2 mixing ratio observations are a key data constraint on biospheric CO_2 fluxes, a reconciliation of top-down flux estimates across inversion studies could be especially useful for evaluating biospheric model estimates and their process-based assumptions.

As a step towards reconciling CO_2 flux estimates from regional inversions and biospheric models, this study presents a regional grid-scale geostatistical inversion (Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008; Gourdji et al., 2010) for North America in 2004, using the sampling network of 9 towers collecting continuous CO_2 measurements at that time, as well as available surface flask and aircraft data. The geostatistical inverse modeling (GIM) 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), and relies strongly on the atmospheric measurements and spatiotemporal flux covariance assumptions to help constrain the problem. 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. Four sets of geostatistical results are presented, with and without process-based auxiliary datasets that can help to explain biospheric fluxes, and using two plausible sets of continental boundary conditions.

Geostatistical inversion 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 the various approaches. Geostatistical inversion estimates are also used to interpret the 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., in review).

2. Data and Methods

The presented inversions use the setup described as part of a synthetic data study for June 2004 (i.e. Gourdji et al., 2010) and further evaluated using synthetic data for a full year. A brief overview of the setup and methods is provided below, with methodological details available in Gourdji et al. (2010).

2.1 Flux domain and resolution

Fluxes are estimated at a 1°x1°, 3-hourly resolution for the North American continent in 2004. Positive values indicate a flux from the biosphere to the atmosphere, and conversely, negative values refer to net uptake. The estimation domain spans from 10°N to 70°N and 50°W to 170°W, yielding 2635 land grid-cells (Figure 1). Altogether, 8 million fluxes are estimated for the year (2635 regions x 366 days x 8 flux periods per day). Resolving fluxes for a 4-day-averaged diurnal cycle temporal resolution, as described in Gourdji et al. (2010), was also

explored to reduce computational expense, but the penalty in terms of the quality of the flux estimates was found to be unacceptable.

Flux estimates are not well-constrained at this fine spatiotemporal estimation resolution by the limited atmospheric 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 aggregating to better-constrained resolutions *a posteriori*. 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., 2011) or assumed flat, rather than estimated directly as in the current study.

2.2 Geostatistical inversions

GIM (e.g. Hoeksema & Kitanidis, 1984; Zimmerman et al., 1998) has been used to identify atmospheric trace gas sources and sinks (e.g. Michalak et al., 2004; Mueller et al., 2008; Gourdji et al., 2008; 2010; Miller et al., in review). Although GIM is Bayesian, it differs from synthesis Bayesian inversions (e.g. Baker et al., 2006; Peters et al., 2007; Butler et al., 2010) in a few key ways. GIM does not rely on a set of explicit prior flux estimates derived from biospheric models, fossil fuel inventories, fire emissions estimates, and/or ocean flux estimates. Traditional Bayesian inversion studies (e.g. Peters et al., 2007; Butler et al., 2010) typically use such explicit priors both to define a first estimate of flux magnitudes and to fix the fine-scale variability in the inversion, while the atmospheric data are used to adjust fluxes only at larger scales. Instead, GIM estimates fluxes directly at fine scales, and relies on *a priori* information on the spatiotemporal covariance between estimated fluxes to help constrain the problem. The parameters describing this covariance structure are optimized directly using the atmospheric measurements. Some recent synthesis Bayesian inversion studies also estimate fluxes at relatively fine spatial scales and rely on spatial covariance assumptions (e.g. Rödenbeck et al., 2003; Carouge et al., 2010a,b; Chevallier et al., 2010; Schuh et al., 2010), with the covariance parameters based on analyses of variability in biospheric models, or sensitivity testing to assess their impact on flux estimates. The GIM estimates presented here also

estimate, rather than prescribe the diurnal cycle based on results from Gourdji et al. (2010) and Huntzinger et al. (2011). We are not aware of any other regional inversions that have taken this step, which is aimed at minimizing temporal aggregation errors that can result from fluxes being adjusted at coarser (e.g. weekly to monthly) scales (e.g. Peters et al., 2007; Butler et al., 2010; Schuh et al., 2010).

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

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

where $\mathbf{z}$ represents the atmospheric CO_2 measurements [ppm], and $\mathbf{s}$ are the unknown surface CO_2 fluxes [µmol/(m^2s)]. $\mathbf{H}$ describes the sensitivity of CO_2 measurements to fluxes, as quantified from an atmospheric transport model, with units of [ppm/(µmol/(m^2s))]. $\mathbf{X}$ contains any pre-selected flux covariates or alternately, one or more columns of ones (see Section 2.4), $\boldsymbol{\beta}$ are the associated drift coefficients, and $\mathbf{X}\boldsymbol{\beta}$ is the component of the flux variability that can be explained by the covariates, a.k.a. the "trend." Fluxes ($\mathbf{s}$) and regression coefficients ($\boldsymbol{\beta}$) are optimized simultaneously by minimizing Eq. 1. $\mathbf{R}$ is the model-data mismatch covariance matrix, describing the expected magnitude of discrepancies between observed ($\mathbf{z}$) and modeled ($\mathbf{Hs}$) CO_2 mixing ratios (due to measurement, transport, representation, and aggregation errors). $\mathbf{Q}$ is the *a priori* flux covariance matrix, characterizing how flux deviations from the model of the trend (i.e. $\mathbf{s}-\mathbf{X}\boldsymbol{\beta}$) are correlated in time and space.

The structure of $\mathbf{Q}$ and $\mathbf{R}$ remain the same as in Gourdji et al. (2010), with a few minor modifications. Separate model-data mismatch values were optimized in $\mathbf{R}$ for each continuous measurement location, as well as for flask and aircraft data (with these latter two representing averages across sites). Additionally, covariance parameters are allowed to vary by month. In particular, spatial flux covariance has been found to have a strong seasonal cycle (Huntzinger et al., 2010) that should be appropriately accounted for in order to yield realistic estimates from the inversion. Monthly model-data mismatch values also help to account for seasonal variations in the quality of the transport model and inversion setup. All covariance parameters are estimated simultaneously, per month, 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).

The solution method for estimating fluxes ($\hat{\mathbf{s}}$), drift coefficients ($\hat{\boldsymbol{\beta}}$) and their *a posteriori* covariances ($V_{\hat{\mathbf{s}}}$ and $V_{\hat{\boldsymbol{\beta}}}$) is described in Gourdji et al. (2010).

2.3 Atmospheric data, boundary conditions and transport

2.3.1 Atmospheric CO_2 mixing ratio measurements

This study uses continuous, high-precision, well-calibrated CO_2 mixing ratio measurements from 9 observational locations unevenly spaced across the North American continent in 2004 (Figure 1). These include two tall towers with a height of 457m (Moody, Texas) and 396m (Park Falls, Wisconsin), two coastal towers less than 25m in height (Sable Island, Nova Scotia and Barrow, Alaska), and five other inland, continental towers ranging in height from 30 to 107m (Norman, Oklahoma; Harvard Forest, Massachusetts; Argyle, Maine; Fraserdale, Ontario; Candle Lake, Saskatchewan). In addition, all available flask and aircraft measurements for this year are included, with the exception of flask samples at coincident tower locations and some coastal sites where the atmospheric transport model was deemed unreliable. Appendix A provides additional details on CO_2 data processing and filtering.

Although the North American measurement network has now expanded to more than 40 sites collecting continuous CO_2 mixing ratio measurements, many new sites are in complex terrain, urban areas, or are very short, and the optimal use of these data in inversions remains a topic of active research (Mueller, 2011; Manning, 2011). This study therefore provides 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_2 concentrations of air flowing into the domain of interest (i.e. the North American land mass here). The impact of these boundary CO_2 mixing ratios on the observations used in the inversion must be pre-subtracted before inferring CO_2 fluxes. Two plausible sets of CO_2 boundary conditions are used in this study: one optimized as part of the CarbonTracker (termed "CT" in this work) global CO_2 data assimilation system (Peters et al., 2007), and the other derived more empirically (termed "EMP" in this work) from marine boundary layer and

aircraft observations taken from the GLOBALVIEW-CO_2 (2010) database. Appendix A provides additional details on these datasets. Only results using the EMP boundary conditions are presented at the monthly timescale, but results with both sets of boundary conditions are discussed at the annual timescale, where they were seen to have a larger impact on the conclusions.

2.3.3 Atmospheric transport model

Surface influence functions ("footprints", or adjoint sensitivities) express the sensitivity of individual CO_2 measurements at specific points in space and time to surface fluxes in the upwind source regions. 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 Appendix A.

The footprints can also be used to assess which portions of the continent are typically constrained by the measurements included in the inversion. Given the limited network in 2004, not all portions of North America are equally well-constrained, as can be seen in the annual average footprint shown in Figure 2a. Not surprisingly, the best-constrained areas are upwind of the measurement locations, in the Central and Northeastern continental United States (U.S.) and a large portion of central Canada. Figure 2b shows three ecoregions across North America with similar land-cover and/or climatic characteristics (modified from Olson et al., 2001) that

are relatively well-constrained by the network (although the Eastern Temperate Forests lack sensitivity in the Southeast U.S.) Inversion results are aggregated *a posteriori to* these three ecoregions for comparison across all presented models.

2.4 Auxiliary variables and variable selection

Geostatistical inversions can estimate fluxes with varying levels of complexity in the flux covariate matrix ($\mathbf{X}$). In addition to potentially improving flux estimates, introducing auxiliary environmental variables as covariates in $\mathbf{X}$ makes it possible to identify significant flux drivers.

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. Therefore, for this study, the Bayes Information Criterion (BIC) (Schwarz, 1978) is used 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., in review) in order to reduce computational expense. Appendix A 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 the flux covariate matrix ($\mathbf{X}$), with no additional information regarding biospheric processes. The fossil fuel inventory is included to help account for the spatial patterns of the fossil fuel emissions, which have a very different structure from the biospheric fluxes. This setup is referred to as the "Simple" inversion, given that the atmospheric observations, fossil fuel inventory and covariance assumptions provide the only constraint on biospheric fluxes in this setup. This inversion is conceptually similar to using a prior of zero in a synthesis Bayesian inversion setup, and provides an independent comparison to process-based, biospheric model output. For the second inversion, we additionally incorporate auxiliary variables from the North American Regional Reanalysis (NARR; Mesinger et al., 2006) to help explain the biospheric signal, which are selected using the combined BIC and Branch-and-Bound algorithm. This case is termed the "NARR" inversion.

All included flux covariates are defined at the resolution at which fluxes are estimated, i.e. 1˚x1˚, 3-hourly, with each variable defined in a single column for all flux locations and time periods. Therefore, the inferred regression coefficients ($\widehat{\boldsymbol{\beta}}$) represent average relationships

over the entire continent and year, with these averages reflecting the portions of the continent within the measurement footprints.

2.4.1 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 2.0 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, 2010).

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*. In both inversions, 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 biosphere models (Marland et al., 2009).

2.4.2 NARR variables

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). Eleven datasets with diurnal variability were considered for inclusion in the inversion, as well as two derived precipitation variables (average

precipitation over the previous 16 and 30 day intervals, Table 2). This superset of NARR variables primarily relates to water fluxes and availability, although 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). However, given that such datasets have a diurnally-varying relationship to CO_2 flux, this would have necessitated the use of diurnally-varying drift coefficients ($\widehat{\boldsymbol{\beta}}$), adding substantial complexity to the trend. Sensitivity tests showed that this setup did not help to improve flux estimates, and thus, these datasets were not included. 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 a measure of vegetative biomass.

2.5 Comparison of inferred fluxes to estimates from other models

Biospheric *a posteriori* flux estimates from the GIM inversions are compared with those from previous inversion studies (CarbonTracker 2009, i.e. 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 biosphere models that participated in the NACP RCIS (Huntzinger et al., in review). 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 available atmospheric measurements for this year are the same 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. (in review).

Model estimates are compared at monthly and annual timescales, and at grid and aggregated ecoregion and continental spatial scales. The analysis of grid-scale patterns helps to visually assess model output, despite high uncertainties at this fine scale, while the comparison

at aggregated spatial scales helps to evaluate the quality of the biospheric models within the inferred 95% uncertainty bounds from the GIM inversions. Net annual flux estimates from all models are less reliable, given that they represent a small residual on a strongly-varying seasonal cycle, although they are particularly important for understanding the carbon budget of North America, and the locations of net sources and sinks within the continent. Overall, this model inter-comparison gives insight into the CO_2 fluxes at various spatiotemporal scales and the strengths and weaknesses of various model formulations. However, an important caveat is that models may still agree with one another for the wrong reasons, e.g. due to systematic errors in the transport models across inversion studies.

3. Results and Discussion

3.1 Optimized covariance parameters

The monthly flux covariance parameters in $\mathbf{Q}$ provide insights into the underlying variability of the flux field, and how this variability changes throughout the year, while the inferred monthly model-data mismatch parameters in $\mathbf{R}$ quantify the ability of the inversion to take advantage of information contained in the CO_2 mixing ratios.

The spatial flux covariance parameters in $\mathbf{Q}$ have a strong seasonal cycle (Figure 3), consistent with that observed in biospheric model estimates (Huntzinger et al., 2011), with the highest variance in July and August and the lowest in the dormant season from November through April. The spatial correlation lengths are shortest during the height of the growing season (~850 and ~600 km respectively for the Simple and NARR inversions in July), with average values across the year of ~1700 and ~1400 km respectively. The temporal correlation range varied from ~5 days in October, a time of sharp seasonal transitions across the continent, to ~50 days in January, with an average value of 22 days throughout the year for both inversions. In the NARR inversion, the variances and correlation lengths of the portion of the flux not explained by the trend $\mathbf{X\beta}$ are reduced relative to the Simple inversion, indicating that the NARR variables explain a portion of the variability in the flux distribution.

The yearly weighted average of the monthly model-data mismatch variances for each measurement site (or rather, the square root of this weighted average, $\sigma_{\mathbf{R}}$, with the weights

defined by the number of data points per month) is shown for the Simple inversion in Figure 3. The tower with the highest average model-data mismatch is Harvard Forest, which is located in a highly-productive forest (Urbanski et al., 2007) about 100 km west of Boston, and even closer to Worcester and Springfield, Massachusetts, while the 1°x1° grid-cell containing this site includes several other small towns and developed areas. The difficulty in matching the data at this tower is likely due to spatial aggregation and representation errors associated with fossil fuel emissions and heterogeneous land cover in the surrounding region. The two towers with the lowest model-data mismatch are BRW and CDL, the two northernmost sites in the domain, where the 1° longitudinal grid-cell size is smaller (see Figure 1). This could point to potential improvements in inversion performance that would be obtained by resolving fluxes at finer spatial scales, perhaps the resolution of the driving winds in the transport model (in this case 40-km or finer for all of North America).

Model-data mismatch can also have significant seasonal variations. For example, the optimized model-data mismatch standard deviation at ARM is 4.7 ppm in April, 0.8 ppm in July, and 0.0004 ppm in September. The high model-data mismatch in April for ARM implies that the full strength of local uptake evident in the measurement data in this month (see Figure A1 in the supplementary material), most likely due to the winter wheat crop planted upwind of the tower (McPherson et al., 2004), is not fully represented by the inversion.

3.2 Auxiliary variables and drift coefficients

3.2.1 Fossil fuel emissions inventory

If the fossil fuel emissions inventory and atmospheric transport model used in this study were perfect, the associated drift coefficient should be approximately one, whereas other values could imply problems with the inversion setup, systematic transport model errors, a lack of atmospheric constraint, covariance between the fossil fuel and biospheric flux signals, and/or errors in the spatiotemporal patterns and magnitudes of emissions in the inventory dataset. The drift coefficients for the Simple inversions were not significantly different from one ($\widehat{\boldsymbol{\beta}} = 0.91$ and 0.95 for the EMP and CT boundary conditions, respectively, and $\sigma_{\widehat{\boldsymbol{\beta}}}$ = 0.06 for both cases), providing support both for the quality of the inventory and the inversion setup. In

addition, this result makes it relatively easy to separate the biospheric and fossil fuel contributions to the total flux *a posteriori*.

3.2.2 Selected NARR variables

The variables selected by the BIC approach (Section 2.4 and Appendix A), and their inferred relationships to CO_2 flux (Table 2) are consistent with process-based understanding of the biospheric carbon cycle. For example, evapotranspiration explains the largest portion of the uptake signal (as indicated by a large negative $\widehat{\boldsymbol{\beta}}$), which is consistent with the physiological relationship 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 up-scaling this value to landscape scales (Anderson et al., 2003).

The positive $\widehat{\boldsymbol{\beta}}$ value associated with specific humidity is consistent with known drivers of heterotrophic respiration (Lloyd & Taylor, 1994; Ise and Moorcroft, 2006), given that this variable, or the mass of water vapor per unit mass of air, scales with both air temperature and surface moisture. The NARR specific humidity and air temperature variables have a correlation coefficient of 0.83 for 2004, and, consequently the $\widehat{\boldsymbol{\beta}}$ uncertainties (derived from $\mathbf{V}_{\widehat{\boldsymbol{\beta}}}$) on these two variables have a strong *a posteriori* anti-correlation (see Table B1 in the supplementary material). This implies that the air temperature variable is primarily helping to correct the signal associated with specific humidity. The positive $\widehat{\boldsymbol{\beta}}$, or 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 selected NARR variables and inferred $\widehat{\boldsymbol{\beta}}$'s are consistent with process-based understanding of biospheric CO_2 flux, not all processes can be easily included in a statistical regression model like the one used here. For example, disturbance (e.g. storms, flooding, insect infestation) may not be fully reflected in the auxiliary variables available for analysis. In

addition, this study only chose to examine diurnally-varying variables from the NARR, and excluded other possible datasets that could help to explain fluxes at daily or weekly timescales (e.g. a fire emission inventory). Third, while evapotranspiration and canopy conductance implicitly incorporate a measure of live biomass, there are no variables within the NARR superset that can directly represent dead substrate availability for heterotrophic respiration, e.g. crop residues after agricultural harvesting or downed woody debris following a storm. Finally, the NARR variables themselves have known biases and 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 flux estimates in GIM through the spatiotemporally-correlated stochastic component. (See equation (5) in Gourdji et al., 2010, and Figure B1 in the supplementary material for an example of the contribution to net grid-scale fluxes in April by each component of the best estimate).

3.3 Seasonal cycle of biospheric fluxes

This section compares monthly biospheric fluxes from the GIM Simple and NARR inversions, aggregated *a posteriori* from the fluxes estimated at the 3-hourly timescale, to results from other inversion studies and from biospheric models participating in the NACP RCIS (Huntzinger et al., in review). The median of the biospheric models is used to represent the current consensus of the terrestrial ecosystem modeling community, although it may still be subject to systematic errors across models. Although the mean has also been shown to have high skill relative to individual models in previous studies (Schwalm et al., 2010), the median is used here to reduce the impact of outliers. Both the mean and median showed better agreement with GIM estimates than any individual biospheric model included in the inter-comparison.

3.3.1 Grid-scale estimates

Grid-scale spatial patterns are similar between the Simple and NARR inversion monthly fluxes (Figure 4), implying that these patterns are primarily informed by the atmospheric observations and spatiotemporal flux correlations, and can therefore be used to evaluate

biospheric model estimates. The NARR auxiliary variables add some realistic heterogeneity to the grid-scale flux estimates relative to the Simple inversion, as seen by a slightly stronger correspondence with the biospheric model median, particularly in under-constrained regions like the Pacific coast.

The grid-scale GIM estimates have a stronger seasonality relative to the biospheric model median, with larger sources in the winter and stronger sinks in the growing season, but the locations with net uptake and release are generally consistent, especially during months with net uptake. The stronger seasonality of the GIM fluxes may be due to errors in the biospheric model median caused by ignoring models that show the strongest sources or sinks in each month. In addition, GIM inversions that account for both the temporal and spatial covariance of fluxes perform better in terms of capturing grid-scale spatial patterns, but in some cases lead to stronger flux variability relative to inversions that account only for spatial covariance.

The GIM estimates exhibit spatial patterns that are more consistent with the biospheric models during the growing, relative to the dormant season. The spatial correlation coefficient between the NARR inversion estimates and the biospheric model median is 0.57 and 0.48 respectively in April and July, compared to only 0.27 and -0.14 in January and October, respectively. Similarly, results for the dormant season show a larger number of biospheric models being significantly different from the GIM results in more areas of the continent (Figure 4, last column), where the significance is defined as a biospheric model being outside the 95% confidence intervals of the GIM NARR estimates.

The greater similarity during the growing season is likely due to, first, both the inversions and biospheric models having greater skill during periods with net uptake. The increased skill during the growing season has previously been documented for biospheric models in model inter-comparison studies using eddy-covariance data (e.g. Schwalm et al., 2010). For inversions, this greater skill is due to the stronger atmospheric signal during this season, thereby making it easier for inversions to constrain fluxes.

Second, errors in the magnitude or spatial patterns of the fossil fuel inventories, or an inability to properly interpret fossil fuel plumes within the inversion, impact the GIM biospheric

fluxes. These errors would be most evident in the dormant season when fossil fuel emissions represent a larger proportion of the total CO_2 flux. For example, some of the strong sources in the GIM results in the south-central portion of the continent in January may be due to representation errors associated with emission plumes at the Moody, Texas tower.

Third, most of the biospheric models in the NACP RCIS do not specifically account for managed agricultural processes (i.e. fertilizer, irrigation, crop harvest and lateral transport, etc.). While this would reduce model skill throughout the year in agricultural areas of the continent (Lokupitiya et al., 2009; Corbin et al., 2010; Huntzinger et al., in review) and in areas with a large consumption of agricultural products (e.g. by human or livestock populations), it would be particularly evident in the presented October fluxes. The strong sources seen in the inversions in this month reflect a strong build-up of CO_2 relative to background air at three towers in heavily agricultural areas (Candle Lake, Saskatchewan, Park Falls, Wisconsin and Norman, Oklahoma; see Figure A1 in the supplementary material), and are most likely associated with the decay of residual biomass after crop harvesting (Johnson et al., 2006). The stronger GIM sinks northwest of the Norman, Oklahoma tower in April and in the agricultural Midwest during July, are also 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).

3.3.2 Ecoregion-scale estimates

Test inversions using synthetic data showed that monthly inversion results at aggregated spatial scales should be reliable in well-constrained areas of the continent (barring substantial transport errors). Therefore, spatially-aggregated GIM results are compared to other inversions and the biospheric models (Figure 5) for the three ecoregions presented in Figure 2b, as well as for the full continent.

At the ecoregion scale, similarly to the grid-scale, inclusion of NARR auxiliary variables is seen to have no significant impact on GIM flux estimates, consistent with results from Gourdji et al. (2008). Yet, the across-model spread for both inversion studies and biospheric models at these aggregated scales is quite large, and generally wider than the GIM confidence intervals

throughout the year, particularly in the Eastern Temperate Forest and the Temperate, Grass, Savannah, Shrub ecoregions. More generally, the spread across inversion studies is smaller than that seen in the biospheric models, thereby supporting the use of the atmospheric data constraint to evaluate biospheric models.

Compared to the other inversion studies, the seasonal cycle of the GIM estimates is within the spread of the other estimates. The spread across inversions is narrower in the better-constrained ecoregions (i.e. Temperate Grass/ Savannah/ Shrub, and Boreal Forest), pointing to the value of an expanded measurement network for reducing the impact of inversion setup choices and assumptions on 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 at progressively larger spatial scales, at least with the limited 2004 measurement network.

In terms of individual inversions, GIM flux estimates are relatively similar to those from CarbonTracker, particularly in the Eastern Temperate and Boreal Forests, despite significantly different estimation resolutions and transport models (Table 1). The strong growing season uptake in the Butler et al. (2010) study relative to the other inversions, particularly in the Boreal and Eastern Temperate Forests, may be due to aggregation errors associated with using measurements from highly productive areas, and extrapolating this signal too strongly across that study's coarser estimation regions (see Table 1). The monthly concentration-averaging intervals and flux estimation timescale used in the Butler et al. (2010) study also support this conclusion. Aggregation errors could also be affecting the CarbonTracker results in the Temperate Grass/ Savannah/ Shrub ecoregion, where this inversion shows stronger peak uptake in July and August relative to the other studies. The Schuh et al. (2010) results show a weaker seasonal cycle and an earlier peak uptake relative to the other inversions in the Eastern Temperate Forests, and Temperate Grass/ Savannah/ Shrub. Among other potential factors, 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 U.S. relative to other biospheric models. This would imply tighter prior flux uncertainties in Schuh et al. (2010) relative to the Butler et al.

(2010) study, where the use of CASA vs. SiB3 as a prior changes flux estimates only slightly at the ecoregion scale.

The spread among the biospheric models is generally larger than that seen across inversion studies, especially in the Temperate Grass/ Savannah/ Shrub ecoregion, which has a relatively high proportion of agricultural land-cover. In this ecoregion, the GIM inversions show a stronger, later peak uptake relative to the majority of the biospheric models, as well as stronger sources in the dormant season, particularly in October. The lack of agreement in this region may again reflect the lack of skill in modeling agricultural processes in the biospheric models (Lokupitiya et al., 2009; Corbin et al., 2010). In the Eastern Temperate Forests, the GIM results fall within the wide biospheric model spread. EC-MOD and MOD17+, the two biospheric models with the strongest growing season uptake in the Eastern Temperate Forests, appear unrealistic when compared to GIM estimates and uncertainties. In the Boreal Forests and at the continental scale, the spread in the biospheric models is narrower and also more consistent with the GIM results. In the Boreal Forests, all models agree reasonably well on the timing of the seasonal cycle, and the biospheric model median falls within the GIM 95% confidence intervals for all months. 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 spatial scales in both bottom-up and top-down models for the seasonal cycle.

3.4 Net annual sources and sinks

Net annual fluxes represent the small residual of a strong seasonal cycle. Therefore, relatively small monthly uncertainties represent a larger proportion of the annual totals. Also, any biases in the inversions or biospheric models may accumulate at longer timescales. In addition, many biospheric models are specifically not designed to capture net annual sources and sinks at all. For these reasons, the annual flux estimates presented here are considered to be less robust than the monthly estimates.

3.4.1 Grid-scale estimates

Although the annual grid-scale GIM flux estimates (Figure 6) show relatively few areas with statistically significant sources or sinks, the spatial patterns are quite consistent between

the two sets of boundary conditions and also with inventory-based estimates of CO_2 flux. The GIM inversions show net annual uptake in the Eastern U.S., the midwestern agricultural areas in the U.S. and Canada, and parts of the continental western U.S., which is consistent with bottom-up inventory estimates from the State of the Carbon Cycle Report (SOCCR; CCSP, 2007) and a recent inventory study by Hayes et al. (in press). Results for these regions are also consistent with Crevoisier et al. (2010), who used an independent carbon budgeting method for North America based on vertical profiles of CO_2 mixing ratios collected from aircraft observations.

The GIM estimates show net annual sources in the southwest U.S., along the Canadian Pacific Coast, and in the Yucatan peninsula of Mexico. The existence of net sources in the Southwest is consistent with Hayes et al. (in press), who hypothesized that livestock and human consumption of agricultural products grown elsewhere could be contributing to sources in these heavily-populated but arid areas, as did Crevoisier et al. (2010). The lack of annual sources in these areas in the majority of the biospheric models (results not shown) is consistent with the fact that many do not explicitly account for the lateral transport of agricultural products. The net sources from the Mexican tropical forests in GIM are also consistent with inventory-based estimates of emissions from land-use change in this region (Cairns et al., 2000; de Jong et al., 2010), while the GIM sources in the Pacific Northwest most likely represent an artifact of the NARR auxiliary variables in this highly under-constrained region.

While inversions with the two boundary condition datasets return similar spatial patterns, magnitudes differ for net annual flux estimates, with stronger sinks and weaker sources evident in almost all regions with the CT relative to the EMP dataset. Most of these differences at the annual timescale are due to a difference in net uptake from March through August, consistent with the seasonally-varying offset between the two sets of boundary conditions (see Figure A2 in the supplementary material).

3.4.2 Ecoregion-scale estimates

At the ecoregion scale (Figure 7), all inversions and biospheric models show a net biospheric uptake in the Eastern Temperate Forests. In this region, the CarbonTracker and Schuh et al. (2010) results fall within the 95% confidence intervals for 3 of the 4 geostatistical

inversions, while the majority of the biospheric models fall within the confidence intervals of the GIM Simple inversion with EMP boundary conditions. The strong uptake of approximately 1PgC/yr seen here in the Butler et al. (2010) results is inconsistent with the other inversions and all but two of the biospheric models, and may reflect the impact of aggregation errors, as discussed previously. In the Boreal Forests, all inversion studies show net uptake, while the biospheric models span both net sources and sinks. CarbonTracker and the Butler et al. (2010) estimates fall within the confidence intervals of the GIM inversions with EMP and CT boundary conditions, respectively, while the Schuh et al. (2010) inversion does not cover this region. In the Temperate Grass/ Savannah/ Shrub ecoregion, the GIM inversions show a more neutral flux relative to the other inversion models and the biospheric models. It is possible that the biospheric models are under-estimating sources in this region due to their lack of skill in modeling residual biomass and the fate of harvested products, as discussed previously, and the other inversions would then also be biased by their use of these same biospheric models as prior flux estimates. As described previously, the GIM sources in these regions in spring and fall are supported by CO_2 mixing ratio time series at the tower locations, as shown in Figure A1 in the supplementary material.

Differences across models in net annual fluxes are most evident at the continental scale (Figure 7). The spread in the continental budget seen here is larger than that for the continental seasonal cycle (Figure 5), although this partially reflects a difference in units between Figure 7 (PgC/yr), which is not normalized by area, and Figure 5 ($\mu mol/m^2s$).

At this coarsest scale of comparison, the GIM results are strongly affected by the choice of boundary conditions (as in Göckede et al., 2010a, which estimated fluxes for the state of Oregon), where the consistent offset between these two datasets leads to an additive effect on flux estimates in both space and time. With the EMP boundary conditions, GIM returns a biospheric flux that is not significantly different from zero, and reduces the net North American sink by approximately 0.7 to 0.9 PgC/yr relative to the GIM inversion with CT boundary conditions. The CarbonTracker (Peters et al., 2007) estimates are statistically consistent with the GIM inversions with CT boundary conditions at the annual continental scale, supporting the conclusion that boundary conditions are the primary control on fluxes at this coarsest scale,

whereas flux resolution, priors, transport model and data choices, have a stronger impact at sub-domain scales. An inventory-based estimate of the North American carbon balance from Hayes et al. (in press) found a net biospheric uptake of 0.33 PgC/yr, which is statistically consistent with the GIM Simple inversion using the EMP boundary conditions, suggesting that these boundary conditions may be more realistic relative to those from CT. The main conclusion, however, is that the uncertainty associated with boundary conditions must be reduced if regional inversions are to be able to accurately pinpoint net biospheric uptake for the entire spatial domain of interest.

4. Conclusions

The primary goal of this study was to perform an inter-comparison of North American CO_2 flux estimates for a single year (2004) across inverse modeling studies and biospheric models, using results from a geostatistical inversion approach as a benchmark. Fluxes were quantified in the GIM inversions at a 1^o x 1^o spatial and 3-hourly temporal resolution, and covariance parameters and process-based auxiliary information were also optimized using the atmospheric data. By estimating fluxes at substantially 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, and allows for the recovery of sub-continental spatial patterns from the atmospheric CO_2 measurements. In addition, avoiding the use of prior flux estimates from biospheric models avoids any biases caused by shortcomings in process-based representations, and allows for a more independent comparison to flux estimates from terrestrial ecosystem models.

The GIM Simple inversion, that included only a fossil fuel inventory as ancillary data, yielded results that support the quality of the inventory as well as the GIM setup and assumptions. Significant regional-scale spatial and temporal variability was recovered, indicating that the information content of atmospheric observations is greater than implied in past studies. The introduction of NARR auxiliary variables into the second, NARR inversion confirmed that flux patterns in the majority of the continent are constrained by atmospheric observations rather than ancillary data, although these did help to extrapolate the signal to

particularly poorly-constrained regions. NARR inversion results are also consistent with process-based understanding of the drivers of CO_2 fluxes from photosynthesis and respiration, with evapotranspiration explaining a substantial portion of the net uptake signal.

For the grid-scale seasonal cycle, the GIM inversions were found to have more consistent spatial patterns with biospheric models during the growing relative to the dormant season. This could be due to 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 ecoregion-scale, comparisons of inferred seasonal cycles across inversion studies pointed to the strong impact of setup on flux estimates, particularly adherence to the bottom-up prior flux estimates and aggregation errors associated with estimating fluxes at coarse-scales relative to the fine-scale variability embedded in the priors. 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. This suggests that an expanded measurement network will further help to reduce the sensitivity of inversion results to setup assumptions, although systematic transport model errors will still remain a concern.

The comparison of GIM results to the biospheric models at the ecoregion-scale pointed to the need for the biospheric models to better account for crop parameterizations, land-management practices, 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, which were not seen in the majority of the biospheric models. The fact that these sources were also absent in some of the examined inversion studies may also point to the impact of these biospheric model limitations on *a posteriori* flux estimates for inversions based on these models.

At the annual timescale, the GIM inversions showed net uptake in the Eastern Temperate and Boreal Forests, but more of a neutral flux in the Temperate Grass/ Savannah/ Shrub. This latter result is inconsistent with other inversion studies and bottom-up models, which may not have properly accounted for the fate of harvested agricultural products and residual biomass. Net annual sources inferred in the Southwest are also consistent with inventory-based estimates of livestock and human consumption of agricultural products in these sparsely-vegetated areas.

While the boundary conditions used as input into regional inversions were found to have only minor impact on recovered grid-scale spatial patterns, they were found to lead to large differences in the GIM flux estimates at the coarsest spatial and temporal scale of comparison, i.e. the annual continental scale. The use of an empirically-derived boundary condition dataset eliminated the North American carbon sink for this year relative to an inversion relying on model output from CarbonTracker for boundary CO_2 concentrations. This result points to the importance of accurately pinpointing the CO_2 concentrations of incoming air for appropriate carbon budgeting at the scale of countries in North America.

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, 2010). 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. The GIM approach also provides an independent comparison to bottom-up model estimates, thereby helping to provide insight into the currently large spread of biospheric model estimates of regional CO_2 flux, while providing a path forward for improving their formulation of processes at regional scales in future work.

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 thank the 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, available at http://carbontracker.noaa.gov.

Collection and processing of CO_2 data from the ARM tower was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy under contract DE-AC02-05CH11231 as part of the Atmospheric Radiation Measurement Program. Other data providers (not listed 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.

The WRF-STILT development at AER has been funded by the National Science Foundation Atmospheric Chemistry Program (grant # ATM-0836153). We thank Steve Wofsy and Christoph Gerbig 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

Ahmadov, R., C. Gerbig, R. Kretschmer, S. Körner, C. Rödenbeck, P. Bousquet, and M. Ramonet (2009), Comparing high resolution WRF-VPRM simulations and two global CO_2 transport models with coastal tower measurements of CO_2, *Biogeosciences*, 6, 807–817.

Anderson, M.C., W.P. Kustas, and J.M. Norman (2003), Upscaling and downscaling, a regional view of the soil-plant-atmosphere continuum, *Agron. J.*, 95:1408-1423.

Andres, R. J., G. Marland, I. Fung, and E. Matthews (1996), A 1°x 1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950 – 1990, *Global Biogeochem. Cycles*, 10(3), 419– 430, doi:10.1029/96GB01523.

Baker, D. F., R. M. Law, K. R. Gurney, P. Rayner, P. Peylin, A. S. Denning, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, K. Masarie, M. Prather, B. Pak, S. Taguchi, and Z. Zhu (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988-2003, *Global Biogeochemical Cycles, 20*(GB1002), doi: 10.1029/2004GB002439.

Bakwin, P. S., P. P. Tans, D. Hurst and C. Zhao (1998), Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program. *Tellus*, 50B, 401-415.

Baldocchi, D. (2008), Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Austr. J. Botany,* 56, 1–26.

Beer, C., M. Reichstein, P. Ciais, G.D. Farquhar, and D. Papale (2007), Mean annual GPP of Europe derived from its water balance, *Geophys. Res. Lett.*, 34, L05401, doi:10.1029/2006GL029006.

Bonan, G. (2008), *Ecological Climatology*, Cambridge University Press, Cambridge, MA, 2nd ed.

Bukovsky, M. S. and D. J. Karoly (2007), A brief evaluation of precipitation from the North American Regional Reanalysis, *J. Hydrometeor., 8,* 837-846.

Butler, M. P., K. J. Davis, A. S. Denning, and S. R. Kawa (2010), Using continental observations in global atmospheric inversions of CO_2: North American carbon sources and sinks, *Tellus 62B*, 550-572, doi: 10.1111/j.1600-0889.2010.00501.x.

Cadule, P., P. Friedlingstein, L. Bopp, S. Sitch, C.D. Jones, P. Ciais, S.L. Piao, and P. Peylin (2010), Benchmarking coupled climate-carbon models against long-term atmospheric CO_2 measurements, *Glob. Biogeochem. Cy.*, 24, GB2016.

Cairns, M.A., P.K. Haggerty, R. Alvarez, B.H.J. de Jong, and I. Olmsted (2000), Tropical Mexico's recent land-use change: a region's contribution to the global carbon cycle, *Ecol. App.*, 10, 5, 1426-1441, doi:10.2307/2641296.

Canadian Greenhouse Gas Measurement Network, http://www.ec.gc.ca/mges-ghgm/Default.asp?lang=En&n=C5F1AC14-1, last access: June 2, 2011.

Canadian Wind Energy Atlas, http://www.windatlas.ca/en/index.php, last access: June 2, 2011.

CarboEurope Atmosphere Database, http://ce-atmosphere.lsce.ipsl.fr/database/index_database.html, last access: June 2, 2011.

Carouge, C., P. Bousquet, P. Peylin, P.J. Rayner, and P. Ciais (2010a), What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes - Part 1: Potential of the network, *Atm. Chem. Phys.*, 10, 6, 3107-3117.

Carouge, C., P. Peylin, P.J. Rayner, P. Bousquet, F. Chevallier, and P. Ciais (2010b), What can we learn from European continuous atmospheric CO_2 measurements to quantify regional fluxes - Part 2: Sensitivity of flux accuracy to inverse setup, *Atm. Chem. Phys.*, 10, 6, 3119-3129.

Chen, S.H. and W.Y. Sun (2002), A one-dimensional time dependent cloud model, *J. Meteorol. Soc. Jpn.*, 80, 1, 99-118.

Chevallier, F., et al. (2010), CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *Journal of Geophysical Research-Atmospheres,* 115, D21307, doi:10.1029/2010JD013887.

Chou, M.D. and M. Suarez (1994), An efficient thermal infrared radiation parameterization for use in general circulation models, Tech Memo 104606, NASA, Washington, DC.

Climate Change Science Program (2007), *The First State of the Carbon Cycle Report (SOCCR): The North American Carbon Budget and Implications for the Global Carbon Cycle.* A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research (King, A.W., L. Dilling, G.P. Zimmerman, D.M. Fairman, R.A. Houghton, G. Marland, A.Z. Rose, & T.J. Wilbanks (eds.)). National Oceanic and Atmospheric Administration, National Climatic Data Center, Asheville, NC, USA, 242 pp.

Corbin, K.D., A.S. Denning, E.Y. Lokupitiya, A.E. Schuh, N.L. Miles, K.J. Davis, S. Richardson, and I.T. Baker (2010), Assessing the impact of crops on regional CO_2 fluxes and atmospheric concentrations, *Tellus B*, 62, 5, Sp. Iss. SI, 521-532.

Crevoisier, C., C. Sweeney, M. Gloor, J. L. Sarmiento, and P. P. Tans (2010), Regional US carbon sinks from three-dimensional atmospheric CO_2 sampling, *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), 18348–18353.

De Jong, B., C. Anaya, O. Masera, M. Olguin, F. Paz, J. Etchevers, R.D. Martinez, G. Guerrero, and C. Balbontin (2010), Greenhouse gas emissions between 1993 and 2002 from land-use change and forestry in Mexico, *For. Ecol. Mgmt.*, 260, 10, 1689-1701, doi:10.1016/j.foreco.2010.08.11.

Deng, F., J.M. Chen, M. Ishizawa, C.W. Yuen, G. Mo, K. Higuchi, D. Chan, and S. Maksyutov (2007), Global monthly CO_2 flux inversion with a focus over North America, *Tellus B*, 59, 2, 179-190.

Draxler, R. R. and G. D. Hess (1998), An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition, *Aust. Meteorol. Mag., 47,* 295-308.

Ek, M.B., K.E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J.D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Enviromental Prediction operational mesoscale Eta model, *J. Geophys. Res.-Atm.*, 108, D22, 8851.

Elliott, D.L., C.G. Holladay, W.R. Barchet, H.P. Foote, and W.F. Sandusky (1986), *Wind Energy Resource Atlas of the United States*, National Renewable Energy Laboratory, Golden, CO, 80401.

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. Enting, I. G., and G. N. Newsam (1990), Atmospheric constituent inversion problems—Implications for base-line monitoring, *J. Atmos. Chem.*, 11(1–2), 69– 87.

Erickson III, D. J., R.T. Mills, J. Gregg, T.J. Blasing, F.M. Hoffman, R.J. Andres, M. Devries, Z. Zhu, and S.R. Kawa (2008), An estimate of monthly global emissions of anthropogenic CO_2: Impact on the seasonal cycle of atmospheric CO_2. *Journal of Geophysical Research-Biogeosciences*, 113**,** G01023, doi:10.1029/2007JG000435.

Geels, C., M. Gloor, P. Ciais, P. Bousquet, P. Peylin, A.T. Vermeulen, R. Dargaville, T. Aalto, J. Brandt, J.H. Christensen, L.M. Frohn, L. Haszpra, U. Karstens, C. Rödenbeck, M. Ramonet, G. Carboni, and R. Santaguida (2007), Comparing atmospheric transport models for future regional inversions over Europe - Part 1: mapping the atmospheric CO_2 signals, *Atm. Chem. Phys.*, 7, 3461-3479.

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, *Journal of Geophysical Research-Atmospheres*, *108*(D24), 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, *Journal of Geophysical Research-Atmospheres*, *108*(D24).

Gerbig, C., S. Körner, & J.C. Lin (2008), Vertical mixing in atmospheric tracer transport models: error characterization and propagation. *Atm. Chem. Phys.*, 8, 591-602.

Gerbig, C., A.J. Dolman, and M. Heimann (2009), On observational and modeling strategies targeted at regional carbon exchange over continents. *Biogeosciences*, 6, 1949-1959.

GLOBALVIEW-CO_2: Cooperative Atmospheric Data Integration Project - Carbon Dioxide (2010). CD-ROM, NOAA ESRL, Boulder, Colorado [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].

Göckede, M., D.P. Turner, A.M. Michalak, D. Vickers, and B.E. Law (2010a), Sensitivity of a sub-regional scale atmospheric inverse CO_2 modeling framework to boundary conditions, *J. Geophys. Res.*, 115, D24112, doi:10.1029/2010JD014443.

Göckede, M., A. M. Michalak, D. Vickers, D. P. Turner, and B. E. Law (2010b), Atmospheric inverse modeling to constrain regional–scale CO_2 budgets at high spatial and temporal resolution, *J. Geophys. Res.*, *115*, D15113, doi:10.1029/2009JD012257.

Gourdji, S.M., K.L. Mueller, K. Schaefer, & 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, V. Yadav, 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,* 6,151-6,167.

Grell, G.A. & D. Devenyi (2002), A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophys. Res. Lett.*, 29, 14, 1693.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller (2009), High resolution fossil fuel combustion CO_2 emission fluxes for the United States. *Environ. Sci. Technol.* **43**, doi:10.1021/es900806c.

Hanna, S. R. (1982), Applications in air pollution modeling. Atmospheric turbulence and air pollution modeling, F. T. M. Nieuwstadt and H. van Dop, eds., D. Reidel, Norwell, Mass., 358 pp.

Hayes, D.J., D.P. Turner, G. Stinson, A.D. McGuire, Y. Wei, T.O. West, L.S. Heath, B. de Jong, B.G. McConkey, R.A. Birdsey, W.A. Kurz, A.R. Jacobson,D.N. Huntzinger, Y. Pan, W.M. Post, and R.B. Cook (in press), Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data, accepted to Global Change Biology, November, 2011.

Hoeksema, R.J. & P.K. Kitanidis (1984), An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resources Research*, 20, 7, 1003-1020.

Hong, S.Y., Y. Noh, and J. Dudhia (2006), A new vertical diffusion package with an explicit treatment of entrainment processes, *Monthly Weather Review*, 134, 9, 2318-2341.

Huntzinger, D.N., S.M. Gourdji, K.L. Mueller & A.M. Michalak (2010), A systematic approach for comparing modeled biospheric carbon fluxes across regional scales, *Biogeosciences*, 8, 1579-1593, doi:10.5194/bgd-8-1579-2011.

Huntzinger, D.N., S.M. Gourdji, K.L. Mueller & A.M. Michalak (2011), The utility of continuous atmospheric measurements for identifying biospheric CO_2 flux variability, *J.Geophys.Res.-Atm.,* doi:10.1029/2010JD015048, in press.

Huntzinger, D.N., W.M. Post, Y. Wei, A.M. Michalak, T.O. West, A.R. Jacobson, I.T. Baker, J.M. Chen, K.J. Davis, D.J. Hayes, F.M. Hoffman, A.K. Jain, S. Liu, A.D. McGuire, R.P. Neilson, B. Poulter, B.M. Raczka, H.Q. Tian, P. Thornton, E. Tomelleri, N. Viovy, J. Xiao, N. Zeng, M. Zhao, and R. Cook (in review), North American Carbon Program (NACP) Regional Interim Synthesis: Terrestrial Biospheric Model Intercomparison, submitted to *Ecological Modeling*, October, 2011.

Ise, T. and P.R. Moorcroft (2006), The global-scale temperature and moisture dependencies of soil organic carbon decomposition: an analysis using a mechanistic decomposition model, *Biogeochemistry*, 80, 3, 217-231.

Johnson, J. M.-F., R.R. Allmaras, and D.C. Reicosky (2006), Estimating source carbon from crop residues, roots and rhizodeposits using the national grain-yield database, *Agron. J.*, 98, 622-636.

Kaminski, T., P.J. Rayner, M. Heimann, and I.G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, 106, 4703-4715.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour Res, 31*(10), 2411-2419.

Kort, E. A., J. Eluszkiewicz, B. B. Stephens, J. B. Miller, C. Gerbig, T. Nehrkorn, B. C. Daube, J. O. Kaplan, S. Houweling, and S. C. Wofsy (2008), Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, *Geophys. Res. Lett., 35,* L18808, doi:10.1029/2008GL034031.

Kort E. A., A. E. Andrews, E. Dlugokencky, C. Sweeney, A. Hirsch, J. Eluszkiewicz, T. Nehrkorn, A. Michalak, B. Stephens, C. Gerbig, J. B. Miller, J. Kaplan, S. Houweling, B. C. Daube, P. Tans, and S. C. Wofsy (2010), Atmospheric constraints on 2004 emissions of methane and nitrous oxide in North America from atmospheric measurements and receptor-oriented modeling framework, *J. Integr. Environ. Sci., 7:2,* 125-133.

Land, A.H., and A.G. Doig (1960), An automatic method of solving discrete programming problems, *Econometrica*, 28, 3, 497-520.

Law, R.M., P.J. Rayner, L.P. Steele, and I.G. Enting (2002), Using high temporal frequency data for CO_2 inversions, *Global Biogeochem Cy*, 4, 1053, doi:10.1029/2001GB001593.

Lin, J. C., C. Gerbig, S. C. Wofsy, B. C. Daube, A. E. Andrews, 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. Gerbig (2005), Accounting for the effect of transport errors on tracer inversions. *Geophysical Research Letters*, 32, L01802, doi:10.1029/2004GL021227.

Lin, Y.L., R.D. Farley, and H.D. Orville (1983), Bulk parameterization of the snow field in a cloud model, *J. Appl. Meteorol.*, 22, 6, 1065-1092.

Lloyd, J., and J. A. Taylor (1994), On the temperature-dependence of soil respiration, *Functional Ecology*, *8*(3), 315-323.

Lokupitiya, E., A.S. Denning, K. Paustian, I. Baker, K. Schaefer, S. Verma, T. Meyers, C.J. Bernacchi, A. Suyker, and M. Fischer (2009), Incorporation of crop phenology in Simple Biosphere Model (SiBCrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, 6, 6, 969-986.

Manning, A.J. (2011), The challenge of estimating regional trace gas emissions from atmospheric observations, *Phil. Trans. R. Soc. A*, 369, 1943-1954, doi:10.1098/rsta.2010.0321.

Markovic, M., C.G. Jones, K. Winger, and D. Paquin (2009), The surface radiation budget over North America: gridded data assessment and evaluation of regional climate models, *Int. J. Climatol.*, 29, 2226-2240, doi:10.1002/joc.1860.

Marland G., K. Hamal, and M. Jonas (2009), How uncertain are estimates of CO_2 emissions?, *J. Ind. Ecol.,* 13, 1, 4-7.

Masarie, K.A. and P.P. Tans (1995), Extension and integration of atmospheric carbon-dioxide data into a globally consistent measurement record, J*. Geophys. Res.-Atm.*, 100, D6, 11593-11610.

Mass, C.F., D. Ovens, K. Westrick, and B.A. Colle (2002), Does increasing horizontal resolution produce more skillful forecasts?, *Bull. Amer. Meteor. Soc.*, 83, 407-430.

McPherson, R.A., D.J. Stensrud, and K.C. Crawford (2004), The impact of Oklahoma's winter wheat belt on the mesoscale environment, *Monthly Weather Review*, 132, 2, 405-421.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi (2006), North American Regional Reanalysis, *Bull. Amer. Meteor. Soc., 87 (3),* 343–360.

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

Miller, S.M., E.A. Kort, A.I. Hirsch, E.J. Dlugokencky, A.E. Andrews, X. Xu, H. Tian, T. Nehrkorn, J. Eluszkiewicz, A.M. Michalak, and S.C. Wofsy (in revision), Regional sources of nitrous oxide over the

United States: seasonal variation and spatial distribution, *J. Geophys. Res. - Atm.*, doi:10.1029/2011JD016951.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res. – Atm.*, 102, D14, 16663-16682.

Mueller, K.L., S.M. Gourdji, & A.M. Michalak (2008), Global monthly-averaged CO_2 fluxes recovered using a geostatistical inverse modeling approach: 1. Results using atmospheric measurements. *J. Geophys. Res.*, 113(D21114), doi:10.1029/2007JD009734.

Mueller, K.L., V. Yadav, P.S. Curtis, C. Vogel, and A.M. Michalak (2010), Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, *Glob. Biogeochem. Cy.*, 24, GB3023, doi:10.1029/2009GB003642.

Mueller, K. L. (2011), *A Data-Driven Multi-Scale Statistical Investigation of Regional Sources and Sinks to Improve Knowledge of Terrestrial Carbon Cycling*. University of Michigan, 2011. United States -- Michigan: *ProQuest Dissertations & Theses A&I.* Web. 13 Dec. 2011.

Nehrkorn, T., J. Eluszkiewicz, S. C. Wofsy, J. C. Lin, C. Gerbig, M. Longo, and S. Freitas (2010), Coupled Weather Research and Forecasting/Stochastic Time-Inverted Lagrangian Transport (WRF-STILT) model, *Meteor. Atmos. Phys., 107 (1-2),* 51–64, doi:10.1007/s00703-010-0068-x.

NOAA-ESRL Global Monitoring Division Tall Tower Network, http://www.esrl.noaa.gov/gmd/ccgg/towers/index.html, last access: June 2, 2011.

Oda, T. & S. Maksyutov (2010), A very high-resolution global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of night-time lights, 1980-2007, *Atm. Chem. Phys. Disc.*, 10, 16307-16344.

Odman, M.T. (1997), A quantitative analysis of numerical diffusion introduced by advection algorithms in air quality models, *Atmos. Env.*, 31, 13, 1933-1940.

Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'Amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y.Kura, J.F. Lamoreux, W.W. Wettengel, P. Hedao, and K.R. Kassem (2001), Terrestrial ecoregions of the world: a new map of life on earth, *Bioscience*, 51(11), 933-938.

Peters, W., A.R. Jacobson, C. Sweeney, A.E. Andrews, T.J. Conway, K. Masarie, J.B. Miller, L.M.P. Bruhwiler, G. Petron, A.I. Hirsch, D.E.J. Worthy, G.R. van der Werf, J.T. Randerson, P.O. Wennberg, M.C. Krol & P.P. Tans (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, PNAS, 104(48).

Peters., W., M.C. Krol, G.R. van der Werf, S. Houweling, C.D. Jones, J. Hughes, K. Schaefer, K.A. Masarie, A.R. Jacobson, J.B. Miller, C.H. Cho, M. Ramonet, M. Schmidt, L. Ciattaglia, F. Apadula, D. Heltai, F.

Meinhardt, A.G. DiSarra, S. Piacentino, D. Sferlazzo, T. Aalto, J. Hatakka, J. Strom, L. Haszpra, H.A.J. Meijer, S. van der Laan, R.E.M. Neubert, A. Jordan, X. Rodo, J.-A. Morgui, A.T. Vermeulen, E. Popa, K. Rozanski, M. Zimnoch, A.C. Manning, M. Leuenberger, C. Uglietti, A.J. Dolman, P. Ciais, M. Heimann, and P.P. Tans (2010a), Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Glob. Ch. Biol.*, 16, 1317-1337, doi:10.1111/j.1365-2486.2009.02078.x.

Peters, W., J.B. Miller, K.M. Schaefer, I. van der Velde, G. van der Werf, A.J. Dolman, N. Carvalhais, and P.P. Tans (2010b), Forest carbon imbalance information improves atmosphere based carbon data assimilation systems, abstract B31F-0378 presented at 2010 Fall Meeting, AGU, San Francisco, CA, 13-17 Dec.

Prather, M.J., X. Zhua, S.E. Strahan, S.D. Steenrod, and J.M. Rodriguez (2008), Quantifying errors in trace species transport modeling, *Proc. Natl. Acad. Sci.*, 105, 50, 19617-19621.

Randerson, J.T., F.M. Hoffman, P.E. Thornton, N.M. Mahowald, K. Lindsay, Y.H. Lee, C.D. Nevison, S.C. Doney, G. Bonan, R. Stockli, C. Covey, S.W. Running, and I.Y. Fung (2009), Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, *Global Change Biology*, 15, 10, 2462-2484.

Rastigejev, Y., R. Park, M.P. Brenner, and D.J. Jacob (2010), Resolving intercontinental pollution plumes in global models of atmospheric transport, *J. Geophys. Res. – Atm.*, 115, D02302, doi:10.1029/2009JD012568.

Riley, W.J., J.T. Randerson, P.N. Foster, and T.J. Lueker (2005), Influence of terrestrial ecosystems and topography on coastal CO_2 measurements: A case study at Trinidad Head, California, *J. Geophys. Res.*, 110, G01005, doi:10.1029/2004JG000007.

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.

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, *Journal of Geophysical Research-Atmospheres,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.

Schwalm, C.R., C.A. Wiliams, K. Schaefer, R. Anderson, M.A. Arain, I. Baker, A. Barr, T.A. Black, G.S. Chen, J.M. Chen, P. Ciais, K.J. Davis, A. Desai, M. Dietze, D. Dragoni, M.L. Fischer, L.B. Flanagan, R. Grant, L.H. Gu, D. Hollinger, R.C. Izaurralde, C. Kucharik, P. Lafleur, B.E. Law, H. Li, Z.P. Li, S.G. Liu, E. Lokupitiya, Y.Q. Luo, S.Y. Ma, H. Margolis, R. Matamala, H. McCaughey, R.K. Monson, W.C. Oechel, C.H. Peng, B. Poulter, D.T. Price, D.M. Ricciuto, W. Riley, A.K. Sahoo, M. Sprintsin, J.F. Sun, H.Q. Tian,

C. Tonitto, H. Verbeeck, and S.B. Verma (2010), A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis, *J. Geophys. Res.-Biogeo.*, 115, G00H05.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.,* 6(2),461–464, doi:10.1214/aos/1176344136.

Skamarock, W. C. and J. B. Klemp (2008), A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, *J. Comp. Phys., 227,* 3465-3485.

Smith, P.C., N. De Noblet-Ducoudre, P. Ciais, P. Peylin, N. Viovy, Y. Meurdesoif, and A. Bondeau (2010), European-wide simulations of croplands using an improved terrestrial biosphere model: phenology and productivity, *J. Geophys. Res.*, 115, G01014, doi:10.1029/2008JG000800.

Stohl, A., (1998), Computation, accuracy and applications of trajectories—a review and bibliography, *Atmos. Environ.*, *32 (6),* 947-966.

Stohl, A., C. Forster, S. Eckhardt, N. Spichtinger, H. Huntrieser, J. Heland, H. Schlager, S. Wilhelm, F. Arnold, O. Cooper (2003), A backward modeling study of intercontinental pollution transport using aircraft measurements, *J. Geophys. Res.*, *108(D12)*, 4370, doi:10.1029/2002JD002862.

Tans, P., and T. J. Conway (2005), Monthly Atmospheric CO_2 Mixing Ratios from the NOAA CMDL Carbon Cycle Cooperative Global Air Sampling Network, 1968-2002., In *Trends: A Compendium of Data on Global Change,* Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Uliasz, M. (1994), Lagrangian particle modeling in mesoscale applications. *Environmental Modeling II,* ed. P. Zannetti, Computational Mechanics Publications, 71-102.

Uliasz, M. and R.A. Pielke (1990), Receptor-oriented Lagrangian-Eulerian model of mesoscale air pollution dispersion, in *Computer techniques in environmental studies*, edited by P. Zannetti, pp. 57-68, Computational Mechanics, Southampton.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, and J.W. Munger (2007), Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, *J. Geophys. Res.*, 112, G02020, doi:10.1029/2006JG000293.

Ward, E. J. (2008), A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools, *Ecol. Modell.,* 211(1–2), 1–10, doi:10.1016/j.ecolmodel.2007.10.030.

Wen, D., J.C. Lin, F. Meng, P.K. Gbor, Z. He, and J.J. Sloan (2011), Quantitative assessment of upstream source influences on total gaseous mercury observations in Ontario, Canada, *Atmos. Chem. Phys.*, 11, 1405-1415.

West, G.L., J. Steenburgh, and W.Y.Y. Cheng (2007), Spurious grid-scale precipitation in the North American Regional Reanalysis, *Monthly Weather Review*, 135, 2168-2183.

Yadav, V., K. L. Mueller, D. Dragoni, A. M. Michalak (2010), A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, *Biogeosciences*, 7, 1445–1487, doi:10.5194/bgd-7-1445-2010.

Yadav, V., K. L. Mueller, and A. Michalak (in review), A Backward Elimination Discrete Optimization Algorithm for Model Selection in Spatio-temporal Regression Models, submitted to *Geographical Analysis*, May 31, 2011.

Yang, W., N.V. Shabanov, D. Huang, W. Wang, R.E. Dickinson, R.R. Nemani, Y. Knyazikhin, and R.B. Myneni (2006), Analysis of leaf area index products from combination of MODIS Terra and Aqua data, *Remote Sensing of Environment*, 104, 3, 297-312.

Zhao, C., A. E. Andrews, L. Bianco, J. Eluszkiewicz, A. Hirsch, C. MacDonald, T. Nehrkorn, and M. L. Fischer (2009), Atmospheric inverse estimates of methane emissions from Central California, *J. Geophys. Res., 114,* D16302, doi:10.1029/2008JD011671.

Zimmerman, D.A., G. de Marsily, C.A. Gotway, M.G. Marietta, C.L. Axness, R.L. Beauheim, R.L. Bras, J. Carrera, G. Dagan, P.B. Davies, D.P. Gallegos, A. Galli, J. Gomez-Hernandez, P. Grindrod, A.L. Gutjahr, P.K. Kitanidis, A.M. Lavenue, D. McLaughlin, S.P. Neuman, B.S. RamaRao, C. Ravenne, and Y. Rubin (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Res. Res.*, 34, 6, 1373-1413.

Tables

Table 1: Comparison of setup and input data across published inversion studies for North America in 2004.

	Domain	Spatial and temporal resolution	A priori flux covariance	Prior - biospheric	Prior – fossil fuels	Winds & transport model
Butler et al. (2010)	Global	10 sub-regions in North America, monthly	None	SiB3 (hourly) / CASA (monthly mean)	Erickson et al. (2008)	GEOS-4/ PCTM
CarbonTracker 2009 (http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/)	Global	25 eco-regions in North America, weekly	Spatial covariance within ecosystem types	CASA-GFEDv2 (3-hourly)	Spatial patterns from EDGAR-4 (http://edgar.jrc.ec.europa.eu/index.php), with additional seasonal cycle	ECMWF/ TM5
Schuh et al. (2010)	North America	1°x1°, weekly	Spatial covariance using fixed correlation length scales	SiB3 (hourly)	Vulcan in continental U.S. (Gurney et al., 2009); Andres et al. (1996) elsewhere	RAMS/ LPDM
GIM – Simple	North America	1°x1°, 3-hourly	Monthly-varying spatial covariance using correlation length scales and variances estimated with atmospheric data	None	Vulcan in continental U.S. (Gurney et al., 2009); CDIAC/ Night Lights dataset elsewhere (Oda & Maksyutov, 2010)	WRF/ STILT
GIM - NARR				Linear trend with 3-hourly NARR variables, calibrated with atmospheric data		

Table 2: Candidate NARR environmental covariates, with associated $\widehat{\boldsymbol{\beta}}$ values for those variables selected using BIC for inclusion in the NARR inversion with the EMP boundary conditions. Auxiliary variables were normalized to zero mean and unit variance, such that $\widehat{\boldsymbol{\beta}}$ values are directly comparable. Variables with dashes ('---') were considered but not selected by the BIC/ Branch & Bound algorithm. Also shown are the coefficients of variation (CV) for the selected variables (i.e. $\left|\sigma_{\widehat{\boldsymbol{\beta}}}/\widehat{\boldsymbol{\beta}}\right|$).

	$\widehat{\boldsymbol{\beta}}$	CV
Canopy conductance	---	---
Downward shortwave radiation	---	---
Evapotranspiration	-1.16	0.05
Precipitation rate	0.28	0.13
Relative humidity	---	---
Specific humidity	0.48	0.12
Soil moisture	---	---
Air temperature (@ 2m)	-0.22	0.25
Plant canopy water content	---	---
Snow depth	---	---
Snow cover (%)	---	---
16-day average precipitation	---	---
30-day average precipitation	---	---

Figure captions

Figure 1: Domains of nested WRF winds, flux estimation grid, and the locations of towers, flask & aircraft measurements used in the inversions. See Table A1 in the supplementary material for a key to the tower names.

Figure 2: a) Sensitivities of all observations used in the inversion (as shown in Table A1 in the supplementary material) to flux estimates, averaged across all 3-hourly flux time periods for the year for each grid-cell location. Black stars represent the nine continuous measurement locations in 2004. b) Three ecoregions (modified from Olson et al., 2001) used for spatial aggregation of flux estimates. These three ecoregions, along with the grey grid-cells, form the domain for the aggregated North American totals in Figures 5 and 7.

Figure 3, a) & b): Optimized covariance parameters using the RML algorithm with the atmospheric measurements, for the GIM inversions with the EMP boundary conditions. a) Monthly flux standard deviations ($\sigma_{\mathbf{Q}}$) for the Simple and NARR inversions. b) Square root of the yearly average model-data mismatch variances ($\sigma_{\mathbf{R}}$) (weighted by the number of data-points in each month) for 9 towers, flask and aircraft data from the Simple inversion. Location information for each measurement code is included in Table A1 in the supplementary material.

Figure 4: Monthly-averaged grid-scale biospheric fluxes for January, April, July and October from the Simple and NARR inversions using the EMP boundary conditions, and the median of biospheric models participating in the NACP RCIS. Also shown are the grid-scale spatial correlation coefficients (ρ) and Root Mean Squared Difference (RMSD; μmol m^{-2} s^{-1}) between the inversion and the biospheric model median fluxes. Please note the different scales for each month. The plots in the right-most column show the percent of individual biospheric models that are outside the 95% confidence intervals of the NARR inversion for each month and location.

Figure 5: Seasonal cycle of monthly-averaged fluxes aggregated to the three well-constrained ecoregions, as well as to the full continent (Figure 2b). GIM fluxes using the EMP boundary conditions are compared to other inversions and individual 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.

Figure 6: Annually-averaged grid-scale biospheric fluxes from the GIM NARR inversions using the EMP and CT boundary conditions. Also shown are significant sources and sinks (at 67% and 95% confidence levels) for each inversion.

Figure 7: Annual biospheric flux estimates spatially aggregated to the three ecoregions and North America (Figure 2b). Results from the Simple and NARR inversions using the EMP and CT

boundary conditions are compared to results from other inversions and the biospheric models participating in the NACP RCIS. GIM results are shown within their 95% confidence intervals.