Measuring fluxes of trace gases at regional scales by Lagrangian observations: Application to the CO_2 Budget and Rectification Airborne (COBRA) study

J. C. Lin,[1] C. Gerbig,[1] S. C. Wofsy,[1] A. E. Andrews,[2] B. C. Daube,[1] C. A. Grainger,[3] B. B. Stephens,[4] P. S. Bakwin,[2] and D. Y. Hollinger[5]

Received 10 March 2004; revised 7 May 2004; accepted 26 May 2004; published 5 August 2004.

[1] We present a general framework for designing and analyzing Lagrangian-type aircraft observations in order to measure surface fluxes of trace gases on regional scales. Lagrangian experiments minimize uncertainties due to advection by measuring tracer concentrations upstream and downstream of the study region, assuring that observed concentration changes represent fluxes within the region. The framework includes (1) a receptor-oriented model of atmospheric transport, including turbulent dispersion, (2) an upstream tracer boundary condition, (3) a surface flux model that predicts the distribution of tracer fluxes in time and space, and (4) a Bayesian inverse analysis that combines a priori information with observations to yield optimal estimates of tracer fluxes by the flux model. We use a receptor-oriented transport model, the Stochastic Time-Inverted Lagrangian Transport (STILT) model, to simulate ensembles of particles representing air parcels transported backward in time from an observation point (receptor), linking receptor concentrations with upstream locations and surface inputs. STILT provides the capability to forecast flight tracks for Lagrangian experiments in the presence of atmospheric shear and dispersion. STILT may be used to forecast flight tracks that sample the upstream tracer boundary condition, or to analyze the data and provide optimized parameters in the surface flux model. We present a case study of regional scale surface CO_2 fluxes using data over the United States obtained in August 2000 in the CO_2 Budget and Rectification Airborne (COBRA-2000) study. STILT forecasts were obtained using the National Centers for Environmental Prediction Eta model to plan the flight tracks. Results from the Bayesian inversion showed large reductions in a priori errors for estimates of daytime ecosystem uptake of CO_2, but constraints on nighttime respiration fluxes were weaker, due to few observations of CO_2 in the nocturnal boundary layer. Derived CO_2 fluxes from the influence-following analysis differed notably from estimates using a conventional one-dimensional budget (''Boundary Layer Budget'') on a typical day, due to time-variable contributions from forests and croplands. A critical examination of uncertainties in the Lagrangian analyses revealed that the largest uncertainties were associated with errors in forecasting the upstream sampling locations and with aggregation of heterogeneous fluxes at the surface. Suggestions for improvements in future experiments are presented. *INDEX TERMS:* 0315 Atmospheric Composition and Structure: Biosphere/atmosphere interactions; 0322 Atmospheric Composition and Structure: Constituent sources and sinks; 0365 Atmospheric Composition and Structure: Troposphere—composition and chemistry; 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; *KEYWORDS:* CO_2 fluxes, Lagrangian experiments, receptor-oriented modeling

Citation: Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, C. A. Grainger, B. B. Stephens, P. S. Bakwin, and D. Y. Hollinger (2004), Measuring fluxes of trace gases at regional scales by Lagrangian observations: Application to the CO_2 Budget and Rectification Airborne (COBRA) study, *J. Geophys. Res.*, *109*, D15304, doi:10.1029/2004JD004754.

[1]Deparment of Earth and Planetary Sciences and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA.

[2]Climate Monitoring and Diagnostics Laboratory, NOAA, Boulder, Colorado, USA.

[3]Department of Atmospheric Sciences, University of North Dakota, Grand Forks, North Dakota, USA.

[4]Atmospheric Technology Division, National Center for Atmospheric Research, Boulder, Colorado, USA.

[5]Forest Service, U.S. Department of Agriculture, Northeast Research Station, Durham, New Hampshire, USA.

0148-0227/04/2004JD004754$09.00

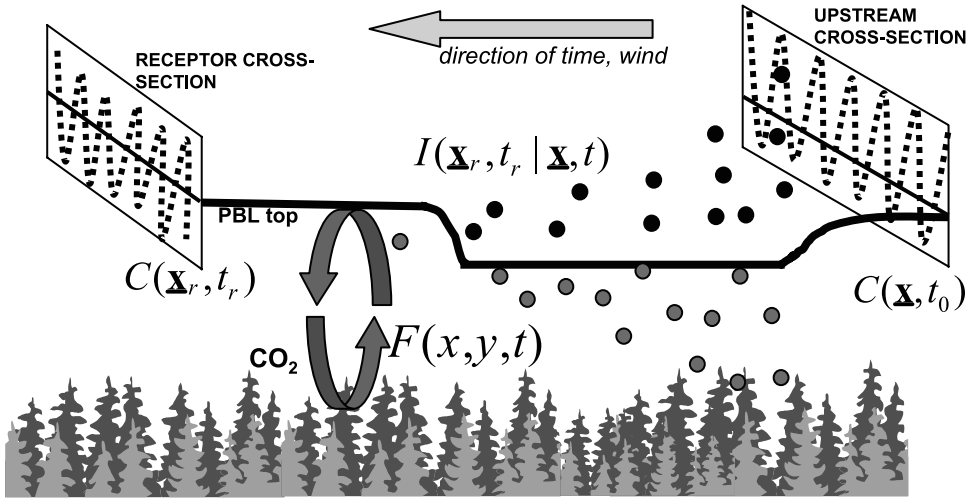

Figure 1. The receptor-oriented analysis framework and the role played by the STILT model. Particle ensembles simulated by STILT provide the influence functions $I(\underline{\mathbf{x}}_r, t_r|\underline{\mathbf{x}}, t)$ that link receptor measurement $C(\underline{\mathbf{x}}_r, t_r)$ to upstream surface fluxes $F(x, y, t)$ and initial tracer field $C(\underline{\mathbf{x}}, t_0)$. The particle ensembles are released at downstream receptors, and their locations prior in time mark out the upstream regions influencing the receptors. To predict upstream sampling locations for Lagrangian experiments, particle locations are calculated in advance using forecasted meteorology. For data analysis the particles are driven with assimilated meteorology to link upstream observations to downstream receptors, quantifying concentration changes that serve as signals from intervening sources/sinks. The simulated concentration changes are quantified using particles that dip into the PBL (shown in grey), which accumulate contributions from surface fluxes generated by a model $F_{mod}(\underline{\mathbf{x}}, t; \underline{\boldsymbol{\lambda}})$ that depends on parameters $\underline{\boldsymbol{\lambda}}$. The analysis framework then uses the deviations between the observed and modeled concentrations to adjust $\underline{\boldsymbol{\lambda}}$ such that the modeled values are optimally consistent with the observed values.

1. Introduction

[2] The process of deriving emergent properties from underlying processes occurring at smaller scales (“upscaling”) represents a “classical conceptual problem in ecology, if not all of science” [*Levin*, 1992]. Budgets of carbon and water at the regional scale (length scale of $10^1 \sim 10^3$ km) cannot be reliably inferred from knowledge of leaf- or tree-level physiology [*Ehleringer and Field*, 1993]. Nevertheless these large-scale budgets are extremely important, representing the core data for managing natural resources [*Newson and Calder*, 1989]. Furthermore, emission fluxes of radiatively and chemically active trace gases to the atmosphere [*Chameides et al.*, 1994; *Crutzen and Ramanathan*, 2000], resulting from the sum of numerous ecological processes and aggregated effects of decisions made by many individual human beings, remain highly uncertain due to errors in upscaling [*Intergovernmental Panel on Climate Change* (*IPCC*), 2001]. Hence there is strong societal motivation to develop methods to use observations to quantify and validate estimates of large-scale CO_2 or other trace gas fluxes derived from scaling up smaller scale processes.

[3] In this paper we discuss a receptor-oriented framework to design and carry out Lagrangian atmospheric experiments and to derive estimates of trace gas fluxes at regional scales (Figure 1). Usually tracer mixing ratios are sensitive to fluxes outside of the target region, and their interpretation is subject to additional uncertainties from these outside fluxes. Lagrangian observations, often conceived as measurements over time and moving with an air mass, restrict surface flux contributions to a limited domain by comparing tracer concentrations measured upstream, at an initial time, and values downstream, at a later time (at the “receptor” location). In this way, the Lagrangian experiments provide tighter, integral constraints on surface fluxes within the target domain and make results insensitive to fluxes outside of the target domain.

[4] Air parcels are typically transported across hundreds of km during a day, undergoing concentration changes due to the intervening influence of surface fluxes. Lagrangian experiments are generally thought to require ideal meteorological conditions with negligible shear and dispersion, a rare circumstance that limits application of the technique [*Schmitgen et al.*, 2004]. The methods developed here relax these constraints and broaden the application of the Lagrangian strategy for determining regional fluxes.

[5] The framework (also see Figure 1 of *Gerbig et al.* [2003b]) consists of (1) a model of atmospheric transport, the Stochastic Time-Inverted Lagrangian Transport (STILT) model [*Lin et al.*, 2003], which simulates transport of air parcels arriving at a receptor, thus linking the receptor with upstream regions; (2) upstream tracer boundary conditions, in this paper directly measured with an aircraft; (3) a surface flux model that predicts the distribution of tracer fluxes in time and space; and (4) a Bayesian inversion that combines prior ground-based flux data with observations from the Lagrangian experiments to adjust parameters of the surface flux model, yielding optimal estimates of surface sources and sinks in the measurement domain. The upstream influences simulated by STILT provide both the information to plan flight tracks for sampling air that will later reach the receptor (forecast mode) and the quantitative link needed to

optimize parameters in the surface flux model (analysis mode). Since STILT simulates the effects of wind shear and dispersion that cause air parcels arriving at a receptor to originate from different air masses, it is technically "influence following" rather than "air mass following."

[6] An aircraft is capable of sampling both the upstream and the downstream in order to carry out Lagrangian experiments. However, aircraft experiments are necessarily constrained by restrictions related to flight safety, logistics, weather, and cost. To be most useful, fluxes derived from Lagrangian experiments must be scaled up to longer time periods to provide regional budgets for trace gases. We use the optimized empirical flux model to "scale up" in time. Data from eddy covariance measurements (e.g., for CO_2 fluxes) are particularly useful in constructing this empirical model, providing detailed, mechanistic information about processes at the surface with comprehensive temporal coverage and high time resolution, but at a limited spatial scale. Thus the framework ingests both aircraft observations and ground-based measurements of concentrations and fluxes in order to optimize parameters in the surface flux model using a Bayesian inverse method. The optimized surface flux model, simultaneously constrained by atmospheric and ground-based data, can then be driven by environmental forcing variables from assimilated meteorological fields in order to extend the fluxes in time to cover periods when parameters of the flux model are deemed to remain steady. This framework is thus an assimilation procedure enabling detailed ground-based information to be "scaled up" to the regional scale by enforcing consistency with large-scale measurements of atmospheric concentration gradients.

[7] We apply the framework to the analysis of regional scale surface CO_2 fluxes from data obtained over the United States in August 2000 as part of the CO_2 Budget and Rectification Airborne (COBRA-2000) study. Current knowledge of CO_2 fluxes at the scale of ecosystems or countries remains highly uncertain [cf. *Schimel et al.*, 2001; *IPCC*, 2001]. Carbon cycle models which incorporate advances in satellite imagery and plant physiology [*Potter et al.*, 1993; *Running and Hunt*, 1993; *Sellers et al.*, 1996] have generated simulations of regional scale carbon fluxes [*Schimel et al.*, 2000], but data to critically evaluate these models at regional scales have been lacking. Continuous eddy covariance measurements on towers have elucidated environmental controls on carbon exchange between the biosphere and atmosphere [*Baldocchi*, 2003; *Goulden et al.*, 1996; *Wofsy et al.*, 1993] at scales of ~1 km, but comprehensive spatial coverage is not possible. "Atmospheric inversion" methods [*Ciais et al.*, 1995; *Fan et al.*, 1998; *Tans et al.*, 1990] combining CO_2 data at remote marine stations with modeled atmospheric transport have characterized carbon fluxes on continental to global scales (10^3~10^4 km) but have yet to yield results at the regional scale due to the dearth of CO_2 observations in proximity to terrestrial sources and sinks [*Sarmiento and Wofsy*, 1999; *Tans et al.*, 1996] and due to difficulties in representing transport processes over the continent in order to interpret the observations [*Gloor et al.*, 1999; *Law et al.*, 1996]. Alternatively, one-dimensional boundary layer budget techniques have been applied to atmospheric CO_2 observations to derive regional scale carbon fluxes [*Denmead et al.*, 1996; *Kuck et al.*, 2000; *Levy et al.*, 1999; *Lloyd et al.*, 2001]. However, horizontal advection neglected in the one-dimensional assumption introduces significant errors that are difficult to account for [*Lin et al.*, 2003; *Cleugh and Grimmond*, 2001].

[8] Aircraft observations using the Lagrangian approach are designed to provide constraints on carbon fluxes at larger spatial scales than the ground-based methods, with greater reliability than conventional boundary layer budgets, addressing the current missing scale in carbon budgets.

[9] We illustrate the application of the analysis framework for planning and analyzing Lagrangian observations—which minimize errors arising from horizontal advection—using data from the CO_2 Budget and Rectification Airborne (COBRA-2000) experiment, a pilot study aimed at testing methods for quantifying regional- and continental-scale fluxes of carbon [*Stephens et al.*, 2000].

[10] In the next section we outline the analysis framework in its general form. In section 3 we adapt the analysis framework for CO_2, presenting COBRA observations and providing details of the surface flux model and the Bayesian optimization. Results of the COBRA analysis are presented in section 4, and an assessment of errors in the analysis and necessary steps to improve current capabilities are presented in section 5. Conclusions derived from this study are shown in section 6.

2. Receptor-Oriented Analysis Framework

2.1. Stochastic Time-Inverted Lagrangian Transport (STILT) Model

[11] We use STILT to simulate the transport of air parcels between the downstream and upstream sampling locations. STILT [*Lin et al.*, 2003] simulates the transport of air parcels with ensembles of representative particles advected with the mean wind, subject to stochastic perturbations parameterized to capture the effects of turbulent transport. The particle ensemble is released at the receptor and transported backward in time, tracing the trajectories of air parcels arriving (in the forward-time sense) at the receptor at a given time.

[12] The density of STILT particles is used to calculate the influence function $I(\underline{\mathbf{x}}_r, t_r|\underline{\mathbf{x}}, t)$ and the footprint $f(\underline{\mathbf{x}}_r, t_r|\underline{\mathbf{x}}, t)$ (see *Lin et al.* [2003] for more details). $I(\underline{\mathbf{x}}_r, t_r|\underline{\mathbf{x}}, t)$ and $f(\underline{\mathbf{x}}_r, t_r|\underline{\mathbf{x}}, t)$ link concentration measurements at the receptor, $C(\underline{\mathbf{x}}_r, t_r)$, to the sum of all upstream contributions:

$$C(\underline{\mathbf{x}}_r, t_r) = \underbrace{\sum_{i,j,m} f\left(\underline{\mathbf{x}}_r, t_r \mid x_i, y_j, t_m\right) \cdot F\left(x_i, y_j, t_m\right)}_{\text{contribution from sources/sinks}} + \underbrace{\sum_{i,j,k} I\left(\underline{\mathbf{x}}_r, t_r \mid x_i, y_j, z_k, t_0\right) \cdot C\left(x_i, y_j, z_k, t_0\right)}_{\text{contribution from advection of upstream tracer field}}, \quad (1)$$

where $F(x_i, y_j, t_m)$ is the surface flux at location (x_i, y_j) and time t_m, and $C(x_i, y_j, z_k, t_0)$ is the initial mixing ratio at time t_0. The first sum on the right-hand side of equation (1) denotes the concentration change at the receptor due to surface fluxes during the time interval between initialization time t_0 and t_r. The second sum refers to the contribution to the receptor concentration from advection of tracers from the initial tracer field $C(x_i, y_j, z_k, t_0)$.

[13] Equation (1) suggests that the initial tracer distribution $C(x_i, y_j, z_k, t_0)$ plays a role only where $I(\underline{\mathbf{x}}_r, t_r|x_i, y_j, z_k, t_0)$ is nonzero, or where particles are found. Therefore $I(\underline{\mathbf{x}}_r, t_r|x_i, y_j, z_k, t_0)$ can be used to forecast optimal locations for sampling the upstream initial tracer mixing ratios. To carry out an influence-following experiment, the upstream influence function $I(\underline{\mathbf{x}}_r, t_r|x_i, y_j, z_k, t_0)$ —a three-dimensional field at each t—must be determined in advance of the measurements in order to plan where and when the aircraft should sample. Section 3.1.2 describes STILT as an operational flight planning tool to determine $I(\underline{\mathbf{x}}_r, t_r|x_i, y_j, z_k, t_0)$.

[14] We now express equation (1) compactly in matrix formulation, in which a single underline denotes a vector and a double underline denotes a matrix:

$$\underline{\mathbf{C}} = \underline{\underline{\mathbf{f}}}\,\underline{\mathbf{F}} + \underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t0}. \tag{2}$$

$\underline{\mathbf{C}}$ is a vector of tracer concentrations at different receptor locations and times. $\underline{\underline{\mathbf{f}}}$ is a matrix of footprint elements relating the receptor concentrations to a vector $\underline{\mathbf{F}}$ of surface fluxes, whose length equals the total number of surface flux elements in the model domain, multiplied by the total number of time steps. $\underline{\underline{\mathbf{I}}}$ is the matrix of influence elements that advects the initial concentration field $\underline{\mathbf{C}}_{t0}$ at time t_0 to the receptors. $\underline{\mathbf{C}}_{t0}$ is a vector with length equal to the total number of gridcells in the model domain.

2.2. Application of Receptor-Oriented Framework to Constrain Tracer Fluxes

[15] Rearrangement of equation (2) illustrates how the observed $\underline{\mathbf{C}}$ and $\underline{\mathbf{C}}_{t0}$ can be quantitatively related to the surface flux $\underline{\mathbf{F}}$:

$$\underbrace{\underline{\mathbf{C}} - \overbrace{\underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t0}}^{\underline{\mathbf{C}}_{up}}}_{\text{Observational constraint}} = \underbrace{\underline{\underline{\mathbf{f}}}\,\underline{\mathbf{F}}}_{\text{Surface flux contribution}}. \tag{3}$$

Equation (3) suggests that knowledge of $\underline{\underline{\mathbf{I}}}$, combined with observations of $\underline{\mathbf{C}}$ and $\underline{\mathbf{C}}_{t0}$, provides spatially integrated constraints on $\underline{\mathbf{F}}$ (Figure 1). We define $\underline{\mathbf{C}}_{up} \equiv \underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t0}$, reflecting the fact that the upstream tracer concentrations advected to the downstream receptors is given by the product of $\underline{\underline{\mathbf{I}}}$ and $\underline{\mathbf{C}}_{t0}$.

2.2.1. Lagrangian Budget-Derived Flux

[16] The observed $\underline{\mathbf{C}}$ and $\underline{\mathbf{C}}_{t0}$, plus information on $\underline{\underline{\mathbf{I}}}$ and $\underline{\underline{\mathbf{f}}}$ from STILT, enables a "Lagrangian budget" that directly provides a footprint-weighted estimate of tracer flux. To show this, we first transform $\underline{\mathbf{C}}$ in order to decrease the variance associated with small-scale vertical gradients typical of the PBL [*Gerbig et al.*, 2003a] by vertically integrating over the receptor altitudes to derive column tracer amounts [*Wofsy et al.*, 1988], represented by $\widetilde{(\cdots)}$ below:

$$\widetilde{(\cdots)}(\underline{\mathbf{x}}_r, t_r) = m_{air}^{-1} \int_{z_{bot}}^{H} (\cdots)(\underline{\mathbf{x}}_r, t_r)\rho(\underline{\mathbf{x}}_r, t_r)dz, \tag{4}$$

where m_{air} is the molar mass of air, ρ is air density, and H is chosen to be just above the maximum PBL height during the day for each receptor. Column amounts are conserved when vertical mixing simply redistributes tracers within the column. This approach reduces errors if, for example, the PBL height is slightly in error.

[17] Each element in the observational constraint $\widetilde{\underline{\mathbf{C}}} - \underline{\underline{\mathbf{I}}}\,\widetilde{\underline{\mathbf{C}}}_{t0}$ derived from equation (3) is in units of [mole/m^2], representing the total column-integrated change in tracer quantity at a location along the downstream cross-section due to fluxes between the upstream and the downstream. Dividing $\widetilde{\underline{\mathbf{C}}} - \underline{\underline{\mathbf{I}}}\,\widetilde{\underline{\mathbf{C}}}_{t0}$ by the elapsed time τ between the downstream and upstream measurements, we derive a vector of fluxes in units of, e.g., [mole/m^2/s]:

$$\frac{\widetilde{\underline{\mathbf{C}}} - \underline{\underline{\mathbf{I}}}\,\widetilde{\underline{\mathbf{C}}}_{t0}}{\tau} = \frac{\widetilde{\underline{\mathbf{C}}} - \widetilde{\underline{\mathbf{C}}}_{up}}{\tau} = \langle \underline{\mathbf{F}} \rangle. \tag{5}$$

We refer to equation (5) as the "Lagrangian budget." After applying $\widetilde{(\cdots)}$ to equation (3) and dividing by τ, we find by comparison to equation (5) that $\langle \underline{\mathbf{F}} \rangle = \underline{\underline{\mathbf{f}}}\,\widetilde{\underline{\mathbf{F}}}/\tau$, suggesting that $\langle \underline{\mathbf{F}} \rangle$ represents a vector of footprint-weighted fluxes.

[18] $\langle \underline{\mathbf{F}} \rangle$ is a direct estimate of the surface tracer flux if the flux is assumed to be invariant within the footprint [*Chou et al.*, 2002]. Alternatively, a model of $\underline{\mathbf{F}}$ can be used to capture the spatiotemporal variability of the flux and optimized as part of a Bayesian inverse analysis, as discussed in the following section.

2.2.2. Flux Model and Bayesian Inverse Analysis

[19] $\underline{\mathbf{F}}$ may be regarded as an implicit function of environmental variables $\underline{\boldsymbol{\gamma}}$—which depend on $\underline{\mathbf{x}}$ and t—that control the surface tracer fluxes (e.g., temperature, population density, vegetation cover, phenology). We incorporate these environmental variables into a surface flux model $\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})$, in which a subset $\underline{\boldsymbol{\lambda}}$ out of $\underline{\boldsymbol{\gamma}}$ are selected as parameters to be optimized in the inverse analysis: $\underline{\mathbf{F}} = \underline{\mathbf{F}}(\underline{\boldsymbol{\gamma}}) \approx \underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})$.

[20] The observed $\underline{\mathbf{C}}$ and $\underline{\mathbf{C}}_{t0}$ can then be related to $\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})$, using equation (3):

$$\underbrace{\underline{\mathbf{C}} - \overbrace{\underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t0}}^{\underline{\mathbf{C}}_{up}}}_{\text{Observational constraint}} = \underbrace{\underline{\underline{\mathbf{f}}}\,\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})}_{\text{Estimate from modeled surface fluxes}} + \underbrace{\underline{\boldsymbol{\varepsilon}}}_{\text{Error}}. \tag{6}$$

The analysis framework uses the observational constraint $\underline{\mathbf{C}} - \underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t0}$ to adjust the model parameters $\boldsymbol{\lambda}$ such that the modeled changes in tracer concentrations are optimally consistent (in a least-square sense) with the observed values.

[21] In the general case, $\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})$ is a nonlinear function of $\underline{\boldsymbol{\lambda}}$, and optimizing the correspondence between modeled and observed $\underline{\mathbf{C}}$ by adjusting $\underline{\boldsymbol{\lambda}}$ requires use of iterative, numerical techniques. However, the optimization problem has a simple analytical solution if $\underline{\mathbf{F}}_{mod}$ is linearly dependent on $\underline{\boldsymbol{\lambda}}$:

$$\begin{aligned}\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}}) &= \lambda_1\underline{\boldsymbol{\varphi}}_1 + \lambda_2\underline{\boldsymbol{\varphi}}_2 + \cdots + \lambda_n\underline{\boldsymbol{\varphi}}_n \\ &= \left[\underline{\boldsymbol{\varphi}}_1\underline{\boldsymbol{\varphi}}_2\cdots\underline{\boldsymbol{\varphi}}_n\right][\lambda_1\lambda_2\cdots\lambda_n]^T \\ &= \underline{\underline{\boldsymbol{\Phi}}}\,\underline{\boldsymbol{\lambda}}.\end{aligned} \tag{7}$$

[22] Substituting $\underline{\underline{\boldsymbol{\Phi}}}\,\underline{\boldsymbol{\lambda}}$ for $\underline{\mathbf{F}}_{mod}(\underline{\boldsymbol{\lambda}})$ in equation (6) results in an equation of the form $\underline{\mathbf{y}} = \underline{\underline{\mathbf{K}}}\,\underline{\boldsymbol{\lambda}} + \underline{\boldsymbol{\varepsilon}}$, where the vector of

observations $\underline{\mathbf{y}}$ is linearly related to the state vector $\underline{\boldsymbol{\lambda}}$ through the Jacobian matrix $\underline{\underline{\mathbf{K}}}$:

$$\underbrace{\underline{\mathbf{C}} - \underline{\underline{\mathbf{I}}}\,\underline{\mathbf{C}}_{t_0}}_{\underline{\mathbf{y}}} = \underbrace{\underline{\mathbf{f}}\,\underline{\underline{\Phi}}}_{\underline{\underline{\mathbf{K}}}}\,\underline{\boldsymbol{\lambda}} + \underset{\text{Error}}{\boldsymbol{\varepsilon}}. \tag{8}$$

[23] The inverse method optimizes the values of the n parameters within the state vector $\underline{\boldsymbol{\lambda}}$. The Bayesian method incorporates prior estimates and their errors in the optimization. We assume that the measurement error $\underline{\mathbf{e}}$ and the errors in $\underline{\boldsymbol{\lambda}}_{prior}$—the prior estimates for $\underline{\boldsymbol{\lambda}}$—are unbiased (mean = 0) and follow Gaussian statistics characterized by error covariance matrices $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ and $\underline{\underline{\mathbf{S}}}_{prior}$, which quantify the degree of constraint provided by the measurements and prior estimates of $\underline{\boldsymbol{\lambda}}$, respectively.

[24] Standard least squares optimization results in posterior estimates for $\underline{\boldsymbol{\lambda}}$ optimally consistent with both the measurements and the prior estimates for gross fluxes, weighted by $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ and $\underline{\underline{\mathbf{S}}}_{prior}$. The estimate of the state vector $\hat{\underline{\boldsymbol{\lambda}}}$ is given by [*Rodgers*, 2000]:

$$\hat{\underline{\boldsymbol{\lambda}}} = \left(\underline{\underline{\mathbf{K}}}^T\underline{\underline{\mathbf{S}}}_{\varepsilon}^{-1}\underline{\underline{\mathbf{K}}} + \underline{\underline{\mathbf{S}}}_{prior}^{-1}\right)^{-1}\left(\underline{\underline{\mathbf{K}}}^T\underline{\underline{\mathbf{S}}}_{\varepsilon}^{-1}\underline{\mathbf{y}} + \underline{\underline{\mathbf{S}}}_{prior}^{-1}\underline{\boldsymbol{\lambda}}_{prior}\right) \tag{9}$$

with the a posteriori error covariance matrix for $\hat{\underline{\boldsymbol{\lambda}}}$ given by $\hat{\underline{\underline{\mathbf{S}}}}_{\lambda} = \left(\underline{\underline{\mathbf{K}}}^T\underline{\underline{\mathbf{S}}}_{\varepsilon}^{-1}\underline{\underline{\mathbf{K}}} + \underline{\underline{\mathbf{S}}}_{prior}^{-1}\right)^{-1}$.

3. Application to Regional-Scale CO_2 Fluxes

[25] We now apply the general receptor-oriented analysis framework developed in section 2 to CO_2. The flux of CO_2 can be separated into contributions from the biosphere and fossil-fuel combustion: $\underline{\mathbf{F}} = \underline{\mathbf{F}}_{veg} + \underline{\mathbf{F}}_{foss}$. The receptor CO_2 concentrations ($\underline{\mathbf{CO}}_2$) can thus be decomposed into contributions due to biospheric fluxes $\Delta\underline{\mathbf{CO}}_{2veg}$, fossil fuel combustion $\Delta\underline{\mathbf{CO}}_{2\ foss}$, and an advected upstream value $\underline{\mathbf{CO}}_{2up}$, and using equation (2):

$$\underline{\mathbf{CO}}_2 = \underbrace{\underline{\mathbf{f}}\,\underline{\underline{\mathbf{F}}}_{veg}\left(\underline{\boldsymbol{\gamma}}\right)}_{\Delta\underline{\mathbf{CO}}_{2veg}} + \underbrace{\underline{\mathbf{f}}\,\underline{\underline{\mathbf{F}}}_{foss}}_{\Delta\underline{\mathbf{CO}}_{2foss}} + \underbrace{\underline{\underline{\mathbf{I}}}\,\underline{\mathbf{CO}}_{2_{t0}}}_{\underline{\mathbf{CO}}_{2up}}. \tag{10}$$

[26] We directly observe $\underline{\mathbf{CO}}_2$ and $\underline{\mathbf{CO}}_{2t0}$ from Lagrangian experiments in the COBRA study (section 3.1) and calculate $\underline{\underline{\mathbf{I}}}$ and $\underline{\mathbf{f}}$ from STILT particles simulated using assimilated meteorology (section 3.2). $\underline{\mathbf{F}}_{foss}$ is derived following the method of *Gerbig et al.* [2003b] and is discussed in section 3.3. We introduce a simple model for $\underline{\mathbf{F}}_{veg}(\underline{\boldsymbol{\gamma}})$ that is linearly dependent on scaling factors that adjust the photosynthetic ($\underline{\mathbf{GEE}}_v$) and respiration ($\underline{\mathbf{R}}_v$) fluxes for each vegetation type v:

$$\begin{aligned}\underline{\mathbf{F}}_{veg}\left(\underline{\boldsymbol{\gamma}}\right) \approx \underline{\mathbf{F}}_{vegmod}(\underline{\boldsymbol{\lambda}}) &= \lambda_1\underline{\boldsymbol{\varphi}}_1 + \lambda_2\underline{\boldsymbol{\varphi}}_2 + \cdots + \lambda_n\underline{\boldsymbol{\varphi}}_n = \underline{\underline{\boldsymbol{\Phi}}}\,\underline{\boldsymbol{\lambda}} \\ &= \lambda_{GEE,v=1}\underline{\mathbf{GEE}}_{v=1} + \lambda_{R,v=1}\underline{\mathbf{R}}_{v=1} \\ &\quad + \lambda_{GEE,v=2}\underline{\mathbf{GEE}}_{v=2} + \lambda_{R,v=2}\underline{\mathbf{R}}_{v=2} + \cdots\end{aligned} \tag{11}$$

$\underline{\mathbf{GEE}}_v$ and $\underline{\mathbf{R}}_v$ are, respectively, functions of downward short-wave radiative flux and temperature (for more details see section 3.4), fitted to biospheric flux observations from AmeriFlux eddy covariance tower sites for different vegetation classes, spatially distributed using the IGBP land surface grid [*Belward et al.*, 1999]. Errors in prior estimates of regional scale carbon fluxes are also derived from the AmeriFlux observations (section 3.4).

[27] We wish to reduce uncertainties in $\underline{\mathbf{F}}_{veg}$ by optimizing $\underline{\mathbf{F}}_{vegmod}$ through the Bayesian method shown in section 2.2.2. We substitute $\underline{\mathbf{F}}_{veg}(\underline{\boldsymbol{\gamma}})$ in equation (10) with $\underline{\mathbf{F}}_{vegmod}(\underline{\boldsymbol{\lambda}})$ and rearrange:

$$\begin{aligned}&\underline{\mathbf{CO}}_2 - \underline{\underline{\mathbf{I}}}\,\underline{\mathbf{CO}}_{2_{t0}} - \underline{\mathbf{f}}\,\underline{\underline{\mathbf{F}}}_{foss} = \underline{\mathbf{f}}\,\underline{\underline{\mathbf{F}}}_{vegmod}(\underline{\boldsymbol{\lambda}}) + \underline{\boldsymbol{\varepsilon}} \\ &\underbrace{\underline{\mathbf{CO}}_2 - \underline{\mathbf{CO}}_{2up} - \Delta\underline{\mathbf{CO}}_{2_{foss}}}_{\Delta\underline{\mathbf{CO}}_{2veg}} = \underbrace{\underline{\mathbf{f}}\,\underline{\underline{\mathbf{F}}}_{vegmod}(\underline{\boldsymbol{\lambda}})}_{\Delta\underline{\mathbf{CO}}_{2vegmod}} + \underline{\boldsymbol{\varepsilon}} \\ &\Delta\underline{\mathbf{CO}}_{2veg} = \underline{\mathbf{f}}\,\underline{\underline{\Phi}}\,\underline{\boldsymbol{\lambda}} + \underline{\boldsymbol{\varepsilon}}\end{aligned} \tag{12}$$

The regional-scale spatial constraint from measuring $\underline{\mathbf{CO}}_2$ and $\underline{\mathbf{CO}}_{2t0}$ in the Lagrangian experiment, the STILT-simulated $\underline{\underline{\mathbf{I}}}$ and $\underline{\mathbf{F}}$, and the prior information about the biosphere incorporated in $\underline{\mathbf{F}}_{vegmod}$, are all combined in the analysis framework as suggested by equation (12) to optimize $\underline{\boldsymbol{\lambda}}$ in the biospheric model.

[28] We integrate $\Delta\underline{\mathbf{CO}}_{2veg}$ through the atmospheric column (equation (4)) and, following the left-hand side of equation (12):

$$\Delta\widetilde{\underline{\mathbf{CO}}}_{2veg} = \widetilde{\underline{\mathbf{CO}}}_2 - \widetilde{\underline{\mathbf{CO}}}_{2up} - \Delta\widetilde{\underline{\mathbf{CO}}}_{2foss}. \tag{13}$$

The observational constraint $\underline{\mathbf{y}}$ consists of the changes in vertically integrated CO_2 amounts attributed to the biosphere, with one element for each receptor j at location $\underline{\mathbf{x}}_{rj}$, and at time t_r: $\underline{\mathbf{y}} = \Delta\widetilde{\underline{\mathbf{CO}}}_{2veg} = [\Delta\widetilde{CO}_{2,veg}(\underline{\mathbf{x}}_{r1}, t_{r1})\cdots\Delta\widetilde{CO}_{2,veg}(\underline{\mathbf{x}}_{rj}, t_{rj})\cdots]^T$. The same vertical integration is applied to $\underline{\mathbf{f}}\,\underline{\underline{\Phi}}$ to form the Jacobian matrix $\underline{\underline{\mathbf{K}}}$, creating the vertically integrated form of the equation $\underline{\mathbf{y}} = \underline{\underline{\mathbf{K}}}\,\underline{\lambda} + \underline{\varepsilon}$ that creates a linear relationship between $\underline{\mathbf{y}}$ and $\underline{\boldsymbol{\lambda}}$:

$$\underbrace{\Delta\widetilde{\underline{\mathbf{CO}}}_{2veg}}_{\underline{\mathbf{y}}} = \underbrace{\widetilde{\underline{\mathbf{f}}\,\underline{\underline{\Phi}}}}_{\underline{\underline{\mathbf{K}}}}\,\underline{\boldsymbol{\lambda}} + \underline{\tilde{\boldsymbol{\varepsilon}}}. \tag{14}$$

We then apply the Bayesian optimization (equation (9)) to optimize $\underline{\boldsymbol{\lambda}}$. The optimized biosphere model, incorporating information from multiple datastreams, can then be forced with meteorological variables driving $\underline{\mathbf{GEE}}_v$ and $\underline{\mathbf{R}}_v$ to generate regional fluxes and trace gases.

3.1. CO_2 Budget and Rectification Airborne (COBRA-2000) Study

[29] The CO_2 Budget and Rectification Airborne (COBRA-2000) study tested the use of Lagrangian experiments to quantify regional- and continental-scale fluxes of CO_2. COBRA collected in situ observations of CO_2, CO, H_2O, and meteorological variables on the University of North Dakota Cessna Citation II for 30 flight legs over the United States in August 2000 (Figure 2). In addition to the Lagrangian experiments, continental-scale flights were conducted for analysis of large-scale fluxes [*Gerbig et al.*, 2003a, 2003b].

3.1.1. Instrumentation

[30] The CO_2 sensor was a modified nondispersive infrared gas analyzer [*Boering et al.*, 1994; *Daube et al.*, 2002] frequently calibrated in-flight with gas mixtures traceable to World Meteorological Organization (WMO) primary stand-

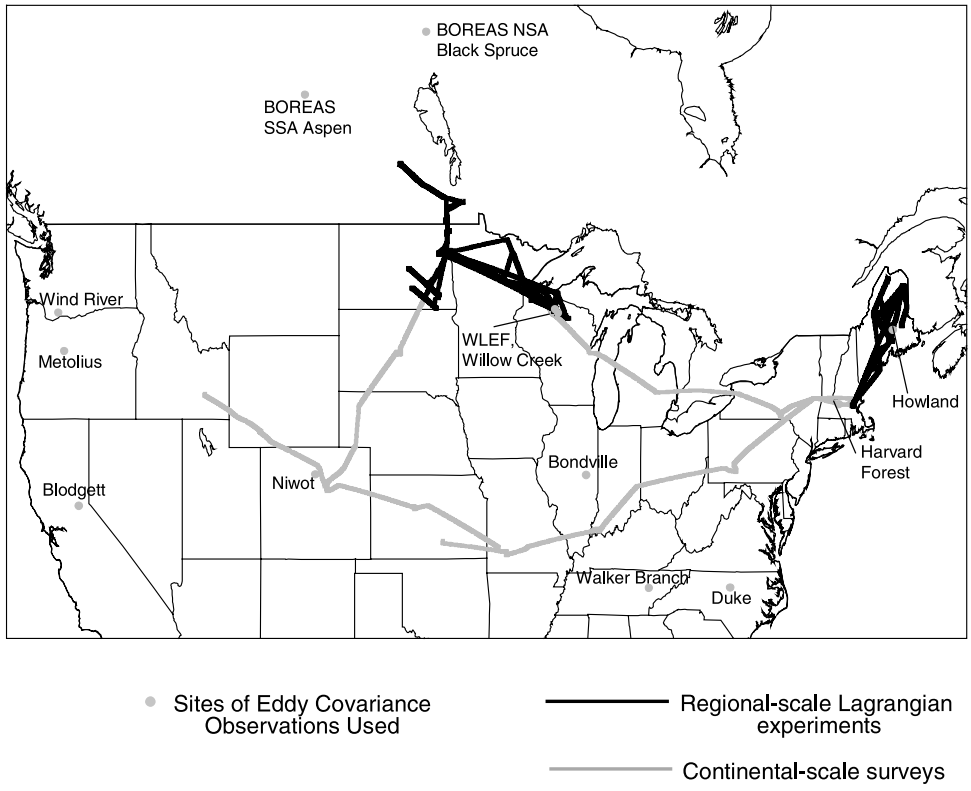

Figure 2. Flight paths conducted by the Cessna Citation II during the CO_2 Budget and Rectification Airborne study (COBRA) and locations of eddy covariance observations from the AmeriFlux network (grey dots) used for generating prior biospheric CO_2 fluxes. The COBRA flights were divided into continental-scale surveys (grey) and the regional scale Lagrangian experiments (black). We examine observations from the COBRA Lagrangian experiments in this paper.

ards [*Conway et al.*, 1994]. Comparison with onboard flask samples and internal "archive" standards indicated uncertainty of the CO_2 observations during COBRA of ±0.25 ppmv [*Daube et al.*, 2002; *Gerbig et al.*, 2003a]. CO_2 mixing ratios were also measured continuously on the WLEF 447-m tall tower in northern Wisconsin at 11, 30, 76, 122, 244, and 396 m above the ground [*Bakwin et al.*, 1998]. These measurements have been ongoing since 1994 and were likewise referenced to the WMO standards. The presence of the WLEF tall tower and its long measurement record provided the motivation to conduct several Lagrangian experiments in northern Wisconsin during COBRA-2000. The CO measurements were acquired using a vacuum-UV resonance fluorescence instrument at 1 Hz resolution with a precision of 2 ppbv and a long-term accuracy of 3 ppbv [*Gerbig et al.*, 1999, 2003a].

3.1.2. Flight Planning in the COBRA Lagrangian Experiments

[31] Upstream influences $\underline{\underline{\mathbf{I}}}$ were predicted using STILT with forecasted winds from the Eta model [*Black*, 1994], and flight tracks were implemented to sample the regions, as illustrated in Figure 3 for receptors in southern ND. Note that the shaded regions represent two-dimensional densities of three-dimensional particle locations projected onto the Earth's surface; some particles are located in the free troposphere and separated by wind shear from the particles in the PBL, e.g., in the long tail of influence stretching to the west in Canada.

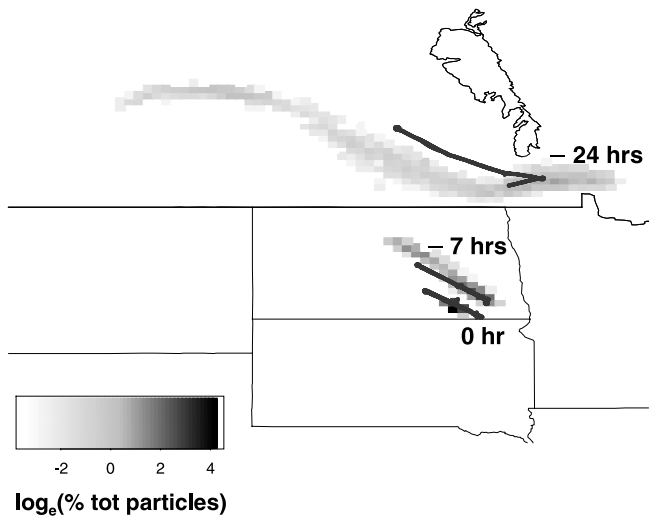

Figure 3. Example of flight planning for Lagrangian experiments. Locations of air parcels at −7 and −24 hours upstream of receptors in southern North Dakota were forecasted by driving the STILT model with forecasted meteorology from the NCEP Eta model. Flight paths were then planned (black lines) in order to sample the particle locations. The greyscale represents particle densities, showing the percentage of the total particles at each time step on a logarithmic scale.

Table 1. Summary of Lagrangian Experiments Conducted as Part of COBRA in August 2000[a]

	Downstream		Upstream		
Name	Date/Time	Location	Date/Time	Location	Experiment Type
ND	2 Aug., UT21	98.56°W, 46.32°N	2 Aug., UT14	98.51°W, 46.85°N	diurnal ($19 \leq \tau \leq 21$)
WI#1	23 Aug., UT22	90.24°W, 46.07°N	23 Aug., UT18	91.62°W, 46.50°N	daytime ($\tau = 4$)
WI#2	23 Aug., UT22	91.10°W, 46.67°N	23 Aug., UT15	92.51°W, 47.33°N	diurnal ($20 \leq \tau \leq 23$)
WI#3	24 Aug., UT22	89.97°W, 45.82°N	24 Aug., UT14	90.65°W, 46.26°N	diurnal ($21 \leq \tau \leq 23$)
WI#4	24 Aug., UT22	89.94°W, 46.20°N	24 Aug., UT14	90.65°W, 46.26°N	diurnal ($21 \leq \tau \leq 23$)
ME	18 Aug., UT19	68.01°W, 46.07°N	18 Aug., UT14	68.70°W, 46.21°N	daytime ($\tau = 5$)

[a]The experiments are separated into diurnal and daytime, depending on whether or not the upstream cross-section sampled the residual layer with tracer signatures remaining from the previous day. Here τ is the number of hours contributing to the observed tracer difference and subject to some uncertainty during the diurnal experiments.

[32] The northerly wind prevalent on 1 August translated into southward shifting Citation flights over the 24 hours in order to sample air parcels arriving at the receptor locations in southern ND on the afternoon of 2 August. The Citation acquired data near Lake Winnipeg in Canada during the afternoon of 1 August and moved southward to central and southern ND during the morning and afternoon of 2 August, respectively, in order to isolate the effect of surface fluxes in ND. During each time period sawtooth flight patterns were conducted to collect numerous vertical profiles at the locations specified by the STILT particle density.

[33] This example illustrates the potential role of a tool like STILT for flight planning: the particles reveal the net effect of turbulent dispersion and wind shear on the spread in air parcel locations, as witnessed in the different hours in this experiment; one cannot easily derive the spread in air parcels from simply examining forecasted wind vectors or mean wind trajectories. Furthermore, the backward-time formulation of STILT yields simulations necessary to derive the upstream influence $\underline{\underline{\mathbf{I}}}$.

[34] Complete sampling of desired particle locations were in some cases limited by logistical considerations: e.g., airspace restrictions, inclement weather conditions, finite flight range, and limited radar coverage. For example, the Citation was not able to dip into the PBL in Canada during the ND flights and could not fully characterize the upstream ($t_0 = -24$ hours) influence. On the Maine flight, a navigation error changed the upstream sampling location, intended for the red square in Figure 9a. Errors in forecasted wind patterns occasionally caused parcels sampled upstream to not arrive at the intended downstream receptors. Sometimes flight patterns could be updated as changes were detected in forecasted wind patterns, allowing relocation of the receptor points to intercept air sampled upstream. Errors arising from the spatial mismatch between the air parcel locations actually sampled versus arriving at the receptor are evaluated and discussed in section 3.5.

3.1.3. Observations

[35] The times and locations of COBRA-2000 Lagrangian experiments in North Dakota (ND), Wisconsin (WI), and Maine (ME) are listed in Table 1, denoted in the discussion below by location and number (e.g., "WI#3"). Receptor observations (denoted $\underline{\mathbf{CO}_2}$) took place during the afternoon, with upstream observations ($\underline{\mathbf{CO}_2}_{\,t0}$) on the morning of the same day except for WI#1, where upstream flights were carried out at noon. Observations and analysis results from each Lagrangian experiment are grouped into separate figures (Figures 4–9), in which panel a shows maps with locations of the upstream and downstream flights as well as results of STILT simulations, panel b displays the tracer observations, panel c shows the simulated vegetation and fossil CO_2 fluxes, and panel d plots the observed CO_2 flux and the total simulated CO_2 fluxes prior to and after Bayesian optimization. The CO_2 fluxes and results from the Bayesian optimization will be discussed in sections 4.2 and 4.3, respectively.

Figure 4. (a) (left) Locations of the upstream (green) and downstream (blue) cross-sections flown by the Citation as part of the ND Lagrangian experiment, as well as the locations of simulated particles (grey and orange) from STILT—started from the downstream cross-section—at the earlier time when the upstream flights were conducted. The particles shown in orange denote those that traveled within the PBL—i.e., particles recently affected by local surface fluxes. The black arrow shows the orientation of the x-axis in the cross-sections shown in Figure 4b, pointing in the direction of increasing x. (right) The time-integrated "footprint" $\underline{\mathbf{f}}$ of the downstream receptors—sensitivity of concentration changes to upstream surface fluxes—derived from particle locations traveling within the PBL shown in orange in the left panel of Figure 4a. The greyscale shows the logarithm (base 10) of the footprint in each 1/6° latitude by 1/4° longitude gridcell. Darker areas denote regions where a unit surface flux leads to a greater change in concentration at the downstream. Note that the left panel shows particles at only a single time when upstream sampling was conducted, but the time-integrated footprint is derived from particle locations at all hours τ separating the times when the downstream and upstream tracer concentrations were affected by surface fluxes. (b) Observed upstream and downstream tracer cross-sections from the ND experiment, showing CO_2, CO, and θ. Flight paths are shown in grey. The origin refers to the mean horizontal position of the aircraft during the sampling of the cross-section. The x-axis represents the horizontal location along the first principal component of the aircraft locations. (c) The modeled CO_2 fluxes attributed to fossil fuel combustion (red dashed), forest (green), and cropland (orange). The modeled results from assuming a maximum value for τ are shown. The error bars are the 1-σ spread resulting from the measurement error $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ (equation (17)). (d) The total modeled biospheric CO_2 flux (black dashed)—sum of the separate components in Figure 4c, the optimized flux after Bayesian inverse analysis (blue dashed), and the observed biospheric flux derived from the Lagrangian budget (solid black; see equation (5)).

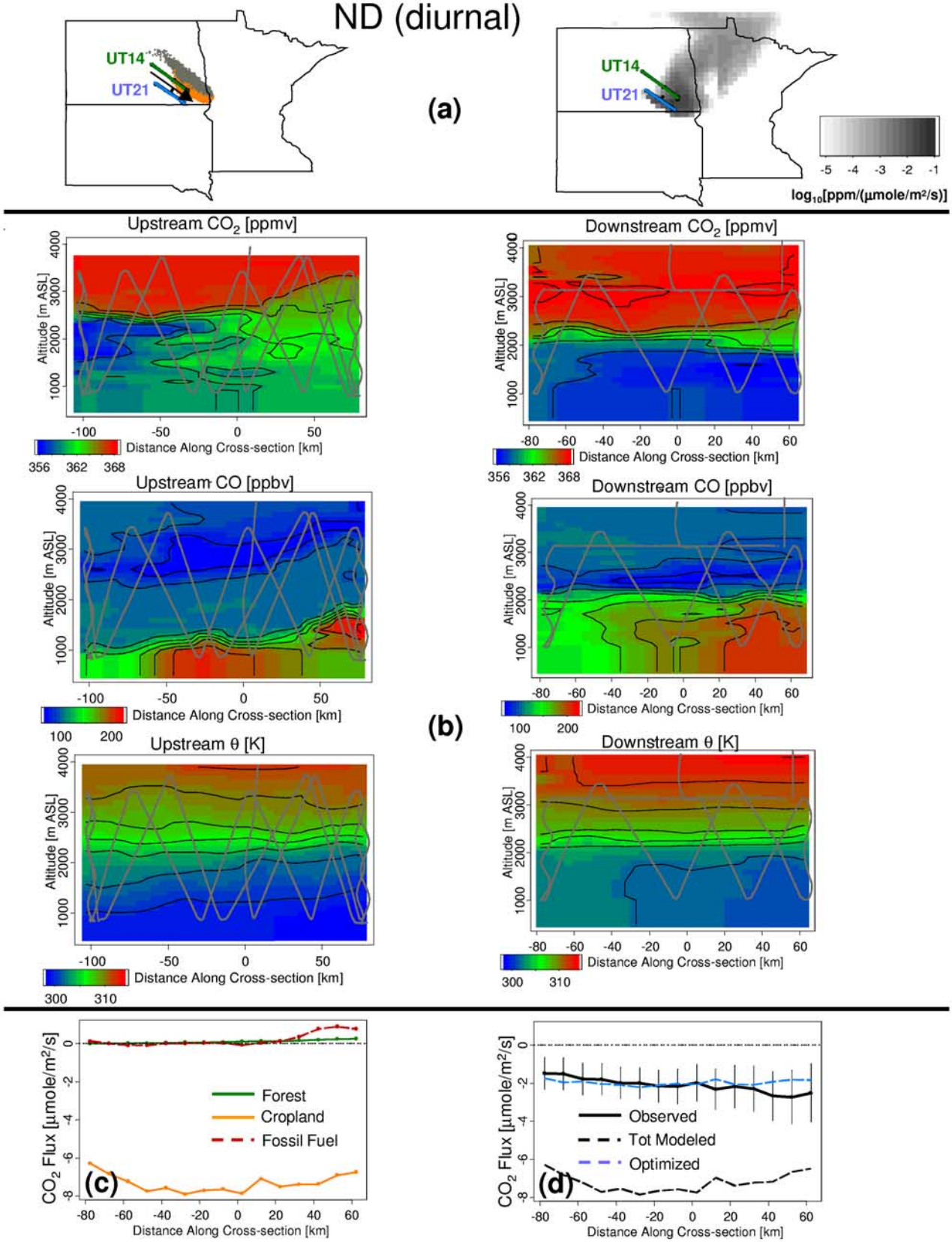
ND (diurnal)
UT14
UT21
(a)
UT14
UT21
-5 -4 -3 -2 -1
log10[ppm/(µmole/m2/s)]
Upstream CO2 [ppmv]
Downstream CO2 [ppmv]
Altitude [m ASL]
Distance Along Cross-section [km]
356 362 368
Upstream CO [ppbv]
Downstream CO [ppbv]
100 200
(b)
Upstream θ [K]
Downstream θ [K]
300 310
CO2 Flux [µmole/m2/s]
Forest
Cropland
Fossil Fuel
(c)
Observed
Tot Modeled
Optimized
(d)

Figure 4

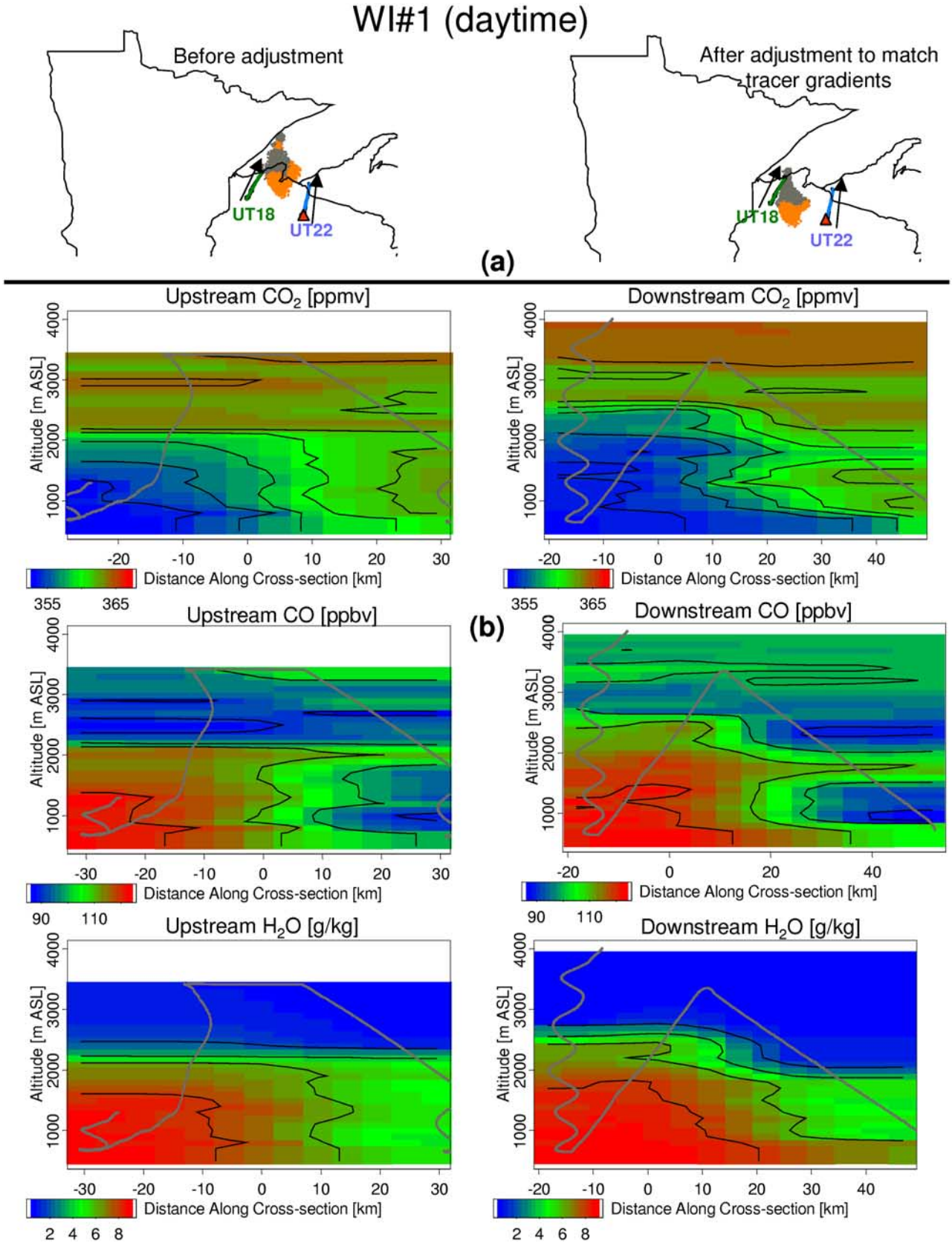

Figure 5. Same as Figure 4, but for the WI#1 experiment. The two panels in Figure 5a show particle distributions (left) before and (right) after adjustment for transport errors (see text). The red triangle shows the location of the WLEF tall tower.

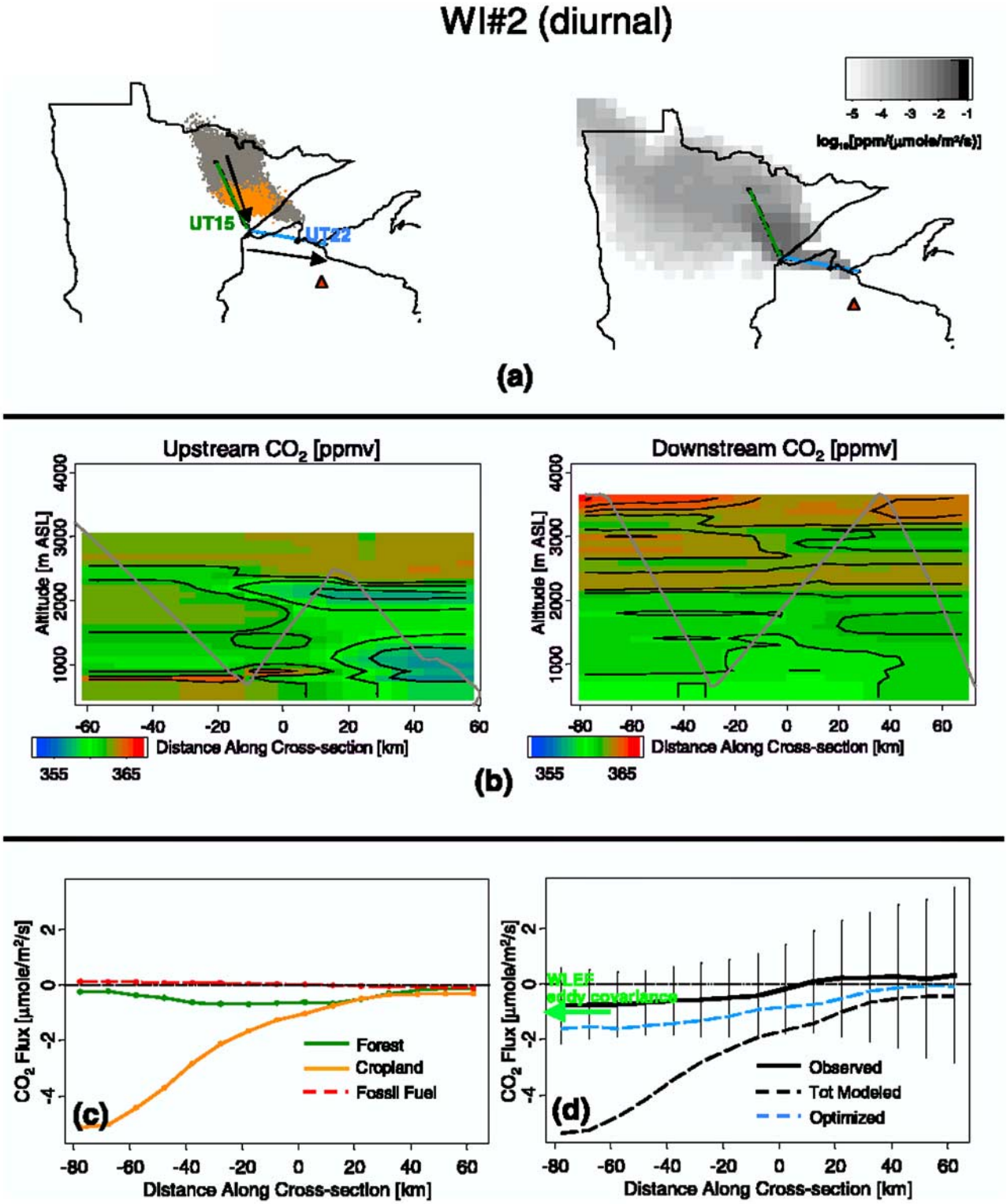

Figure 6. Same as Figure 4, but for the WI#2 experiment. The arrow in light green in Figure 6d refers to the observed CO_2 flux from eddy covariance at WLEF for the same period between 22 and 23 August.

[36] We interpolate observations from the sawtooth flight patterns using an inverse distance squared weighting method to generate continuous tracer cross-sections that facilitate visualization and analysis. We apply subsequent analyses to the continuous tracer cross-sections provided by the interpolation. The tracer cross-sections from the Lagrangian experiments are shown in Figures 4b–9b, with flight paths in grey lines. The *x*-axis refers to the distance along the direction that explains the most variance (the first principal component) in the aircraft's horizontal coordinates, and the origin of the *x*-axis refers to the mean horizontal position of the aircraft path during sampling.

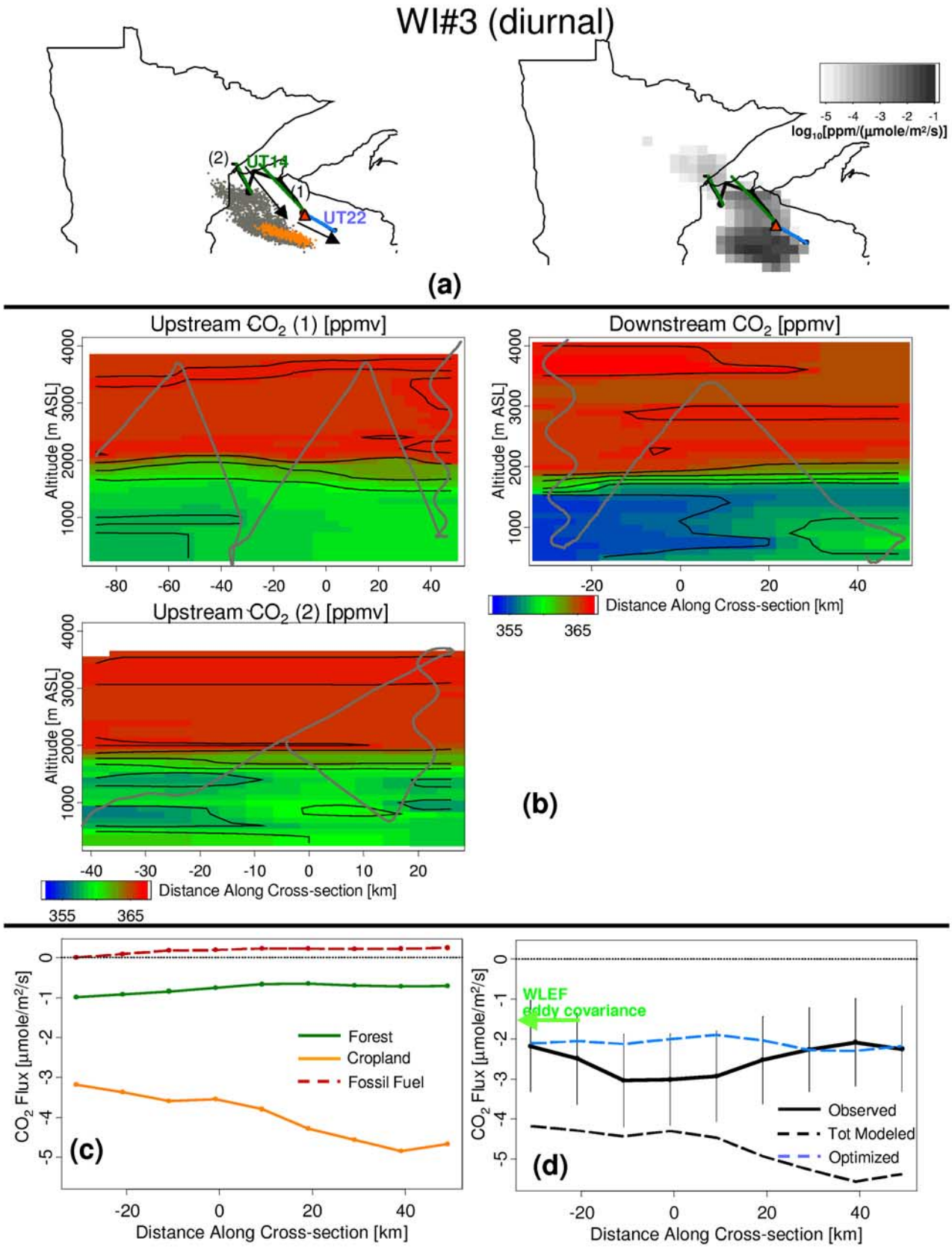

Figure 7. Same as Figure 4, but for the WI#3 experiment. The labels (1) and (2) denote the two separate upstream cross-sections shown in Figure 7b. The arrow in light green in Figure 7d refers to the observed CO_2 flux from eddy covariance at WLEF for the same period between 23 and 24 August.

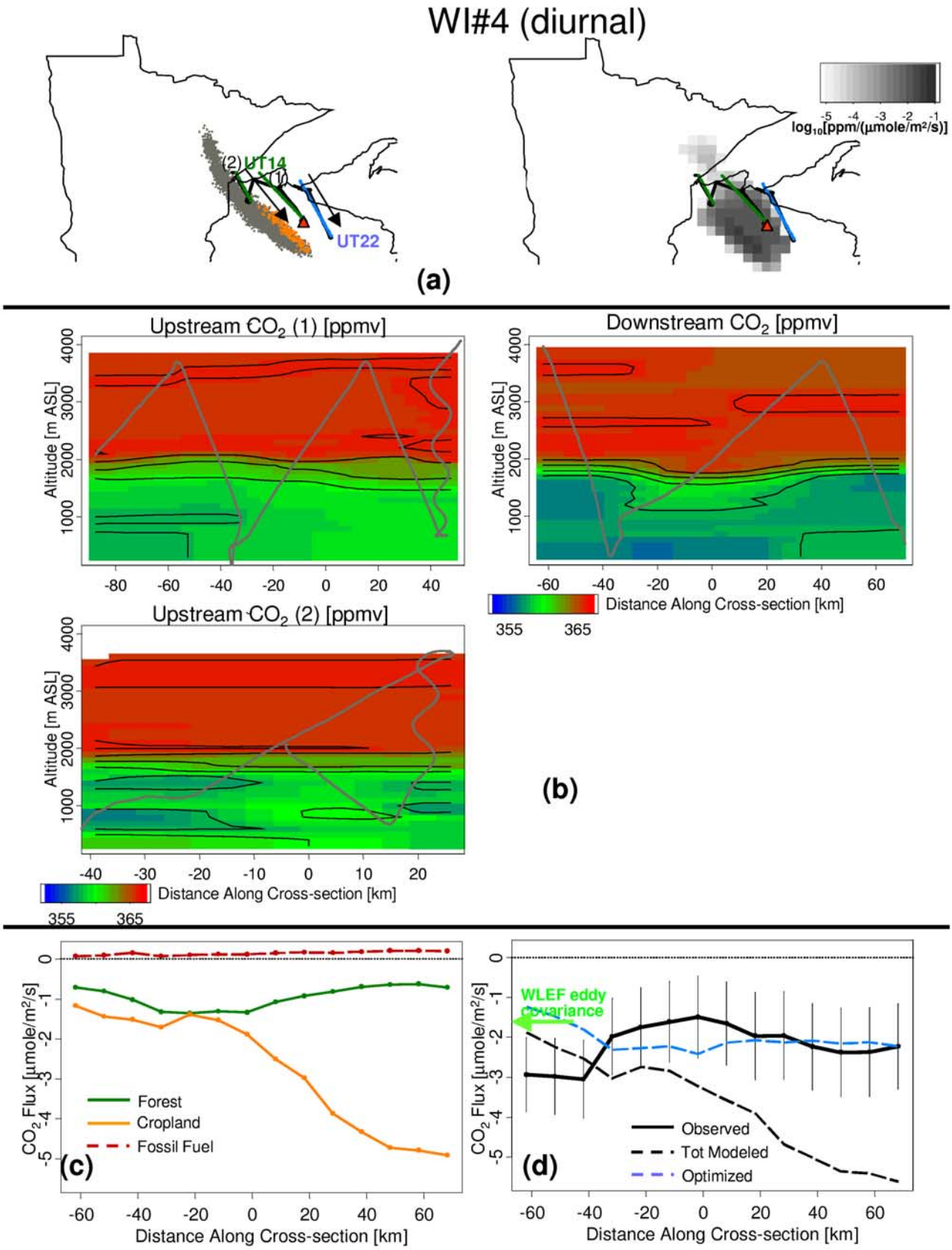

Figure 8. Same as Figure 4, but for the WI#4 experiment. The labels (1) and (2) denote the two separate upstream cross-sections shown in Figure 8b. The arrow in light green in Figure 8d refers to the observed CO_2 flux from eddy covariance at WLEF for the same period between 23 and 24 August.

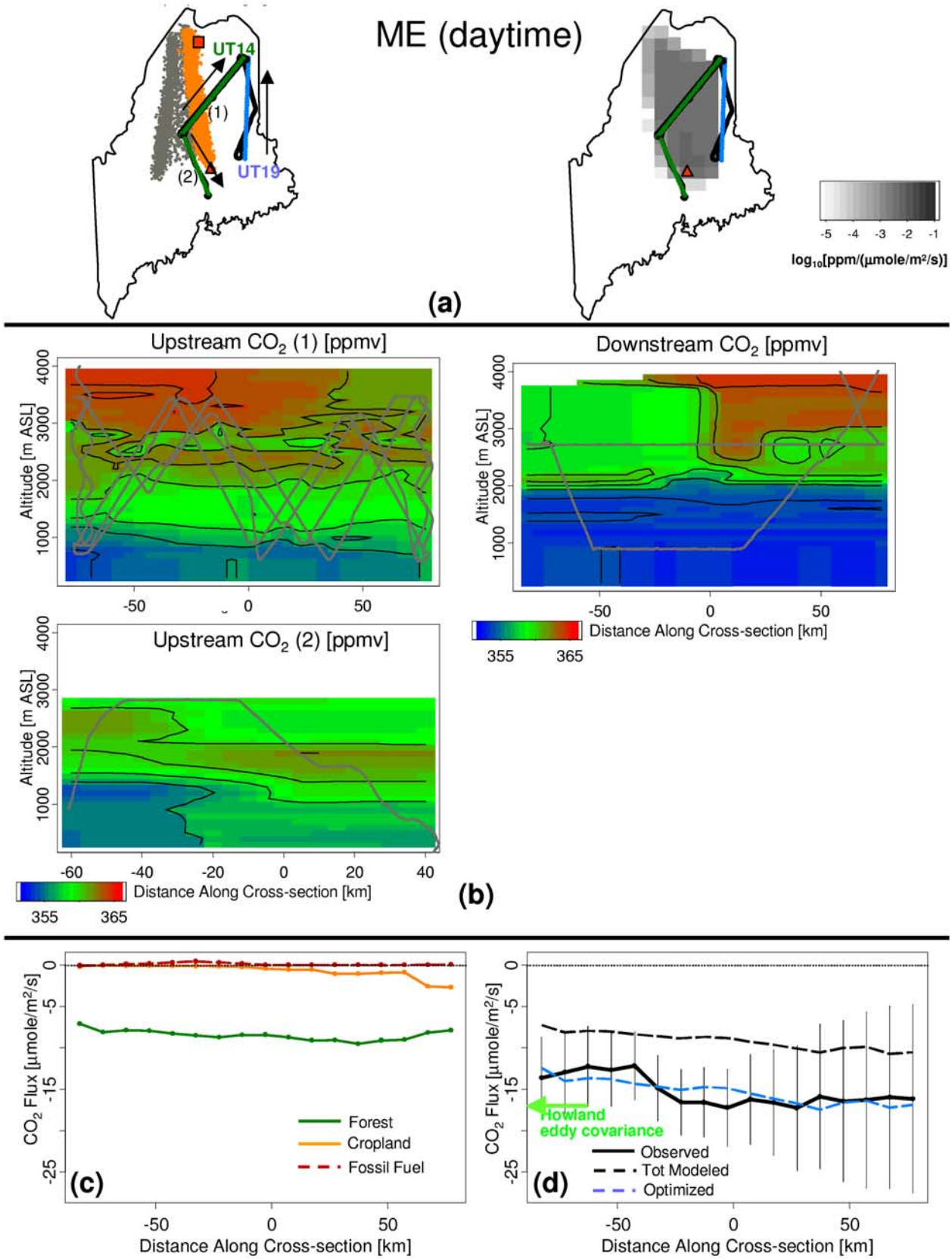

Figure 9. Same as Figure 4, but for the ME experiment. The red triangle in Figure 9a shows the location of the Howland eddy covariance tower. The red square in Figure 9a denotes the original planned flight location for the end of the upstream cross-section (see text). The labels (1) and (2) refer to the two separate upstream cross-sections shown in Figure 9b. The arrow in light green in Figure 9d refers to the observed CO_2 flux from eddy covariance at Howland for the same period (daytime of 18 August).

Figures 4a–9a show the locations of the cross-sections; the orientations of the cross-sections are labeled as black arrows that point in the direction of increasing x in Figures 4b–9b.

[37] The ND experiment (Figure 4) illustrates general features of the observations. The upstream concentrations $\underline{\mathbf{CO}}_{2\ t0}$ in the morning (Figure 4b) show ~4 ppm less CO_2 in the lower altitudes, up to ~2 km, with well-mixed profiles of potential temperature (θ) and elevated values of H_2O (not shown) up to the same altitudes. From these tracer signatures we concluded that the morning cross-section in ND mainly characterized the residual boundary layer, showing well-mixed tracer signatures up to the maximum PBL height during the previous day that remained unchanged after cessation of vertical mixing in the late afternoon. The Citation dipped into the shallow morning mixed-layer on only a few occasions, evidenced by excursions of higher CO_2, CO, θ, and H_2O in the lowest altitudes below 1000 m. By afternoon, the mixed-layer had grown to an altitude similar to the prior residual layer. Depletion of CO_2 at the lower altitudes had grown to ~12 ppm for flights at the receptor.

[38] The CO cross-sections exhibit significantly elevated concentrations, over 200 ppbv in the lower atmosphere, reflecting emissions from forest fires east of Lake Winnipeg on 30 July. The distinctive CO label imparted by the fires, sampled both upstream and downstream, support the STILT analysis indicating that the same general air mass was sampled by the aircraft upstream and downstream on this day. Furthermore, the low CO layer observed at ~3 km during the morning subsided and was resampled again in the afternoon at ~2.5 km. These tracer-derived diagnostics lend confidence to the STILT simulations. We will further use CO as a tracer of combustion to quantify fossil fuel CO_2 emissions (section 3.3) and isolate biospheric contributions to CO_2 changes.

[39] The morning observations in WI#2, WI#3, and WI#4 profiled the residual layer, similar to the ND example. We took care to identify the few observations that dipped into the new morning mixed-layer, as indicated by sharp changes in the other continuously observed tracers (CO, H_2O, and θ). Since these few observations could not be confirmed to be representative of the entire morning mixed-layer, we cannot characterize nighttime or early morning concentrations near the surface, and we excluded the sporadic data in this layer from the analysis. Hence tracer concentrations in the morning reflect values in the residual mixed-layer from the previous day, and column-integrated tracer changes observed between the two cross-sections can then be attributed to the time-integrated fluxes between afternoon of the previous day and downstream (receptor) observations the following afternoon. We thus refer to ND, WI#2, WI#3, and WI#4 as "diurnal experiments" (Table 1), making small corrections for the fact that τ is slightly less than 24 hours, between the time that air in the residual layer was last affected by surface fluxes and when the afternoon mixed-layer was sampled on the following day. This analysis relies on the assumption that the transport of tracers into the residual layer at night may be neglected.

[40] These "diurnal" experiments are contrasted with the "daytime experiments" WI#1 and ME, in which representative observations of concentrations in the PBL were available in the upstream cross-sections (Figure 5b and Figure 9b). The WI#1 upstream flights took place at noon, when the PBL had already grown to altitudes accessible by aircraft. Upstream observations in ME took place at a later time in the morning when vigorous mixing was available, as confirmed by other in situ tracers (not shown). Differences between upstream and downstream concentrations in this case reflect only daytime fluxes.

[41] Large horizontal gradients in CO_2 were observed in the ND and WI#1 experiments, reflecting the spatial heterogeneity in upstream source/sink distributions [*Gerbig et al.*, 2003a]. The marked WI#1 gradient was observed in CO as well as H_2O (Figure 5b). The gradient in WI#1 coincided with land-water contrasts: the left part of the cross-section was more inland, while the right portion was closer to Lake Superior. The air closer to the lake exhibited higher CO_2, lower CO, and lower H_2O.

[42] We were able to carry out one nighttime Lagrangian experiment using the WLEF CO_2 observations as the downstream receptor. We used the aircraft to sample air 9 hours upstream from the nighttime observations at WLEF on 24 August, starting near Lake Superior during the previous afternoon. The overnight buildup of CO_2 led to elevated concentrations of 530 and 383 ppmv at the 11 and 30 m levels on the tower, respectively, notably higher than the 360 ppmv observed aloft. These observations were suitable to constrain nighttime respiratory fluxes of CO_2.

3.2. Particle Simulations Using STILT

[43] To generate the influence $\underline{\underline{\mathbf{I}}}$ and footprint $\underline{\underline{\mathbf{f}}}$ for analysis of the COBRA observations, STILT was driven with assimilated meteorology from the Eta Data Assimilation System (EDAS) [*Rogers et al.*, 1995]. EDAS data, available every 3 hours runs on a 32 km, 45 level grid, is archived by the NOAA Air Resources Laboratory at 80 km horizontal resolution and 22 vertical levels (see http://www.arl.noaa.gov/ss/transport/archives.html).

[44] Particles were released in STILT at receptor points located every 10 km in the horizontal and 200 m in the vertical over the entire downstream cross-section, using 100 particles for each receptor. The column integrals are calculated at each 10 km in the horizontal by vertically integrating the receptor concentrations (equation (4)) available every 200 m up to H; thus $\underline{\widetilde{\mathbf{CO}}}_2$, $\underline{\widetilde{\mathbf{CO}}}_{2up}$, $\Delta\underline{\widetilde{\mathbf{CO}}}_{2foss}$, and $\underline{\widetilde{\mathbf{CO}}}_{2veg}$ are comprised of receptors every 10 km along the downstream cross-section. z_{bot} was chosen to lie below the lowest altitude of the aircraft profiles (~500 m ASL). Where there were systematic errors between modeled and observed PBL heights, we adjusted the PBL heights in the STILT model to match the tracer-derived heights, thereby modifying the vertical extent of turbulent dispersion affecting transport of the particles. This adjustment was applied for the ND and WI#1 experiments. Maximum adjustments were ~400 m during the afternoon.

[45] The resulting footprint $\underline{\underline{\mathbf{f}}}$ will be shown and discussed in section 4.1.2. The matching between $\underline{\underline{\mathbf{I}}}$ and $\underline{\mathbf{CO}}_{2t0}$ to generate $\underline{\mathbf{CO}}_{2up}$ (equation (11)), the advected upstream tracer concentrations, is as follows. $\underline{\mathbf{CO}}_{2\ t0}$ was taken from tracer concentrations at locations in the upstream cross-section (created from interpolation between observations) closest to the particles comprising $\underline{\underline{\mathbf{I}}}$. We will conservatively estimate the error in $\underline{\mathbf{CO}}_{2up}$ associated with the distance

between a particle arriving at the receptor and the upstream data points (see section 3.5).

3.3. Surface Fossil Fuel Emissions

[46] Three surface flux grids—modeled biospheric CO_2 fluxes, fossil CO_2 emissions, and CO emissions—were used to compute $\Delta\underline{\mathbf{CO}}_{2_{vegmod}}$, $\Delta\underline{\mathbf{CO}}_{2_{foss}}$, and $\Delta\underline{\mathbf{CO}}$, respectively. $\Delta\widetilde{CO}_{2_{foss}}(\underline{\mathbf{x}}_r, t_r)$ was estimated from multiplying the Lagrangian budget-derived CO flux $\langle F_{CO}\rangle$ (equation (5)) by the ratio of fossil CO_2:CO enhancements at receptor ($\underline{\mathbf{x}}_r$, t_r) from emission inventories for North America:

$$\Delta\widetilde{CO}_{2foss}(\underline{\mathbf{x}}_r, t_r) = \left.\frac{\delta\widetilde{CO}_{2foss,grid}}{\delta\widetilde{CO}_{grid}}\right|_{\mathbf{x}_r,t_r} \cdot \langle F_{CO}\rangle$$

$$\text{where } \left.\frac{\delta\widetilde{CO}_{2foss,grid}}{\delta\widetilde{CO}_{grid}}\right|_{\mathbf{x}_r,t_r} = \frac{m_{air}^{-1}\int_{z_{bot}}^{H}\rho(\underline{\mathbf{x}}_r,t_r)dz\sum_{i,j,m} f\left(\underline{\mathbf{x}}_r,t_r \mid x_i,y_j,t_m\right)\cdot F_{foss,grid}\left(x_i,y_j,t_m\right)}{m_{air}^{-1}\int_{z_{bot}}^{H}\rho(\underline{\mathbf{x}}_r,t_r)dz\sum_{i,j,m} f\left(\underline{\mathbf{x}}_r,t_r \mid x_i,y_j,t_m\right)\cdot F_{CO,grid}\left(x_i,y_j,t_m\right)} \quad (15)$$

$F_{CO,grid}$ comes from combining the NAPAP 1990 inventory for the northeastern United States (1/6°Lat. × 1/4°Lon.) [*Environmental Protection Agency* (*EPA*), 1993] and the GEIA inventory (1°Lat. × 1°Lon.) [*Benkovitz et al.*, 1996], with hour-of-day and day-of-week scaling factors applied [*Gerbig et al.*, 2003b]. $F_{foss,grid}$ comes from the 1° × 1° inventory compiled by *Marland et al.* [1997], adjusted for increases between 1995 and 2000 as discussed by *Gerbig et al.* [2003b].

[47] *Gerbig et al.* [2003b] adopted the indirect approach of equation (15) to reduce the sensitivity to transport errors, using observed changes in CO instead of directly using the fossil CO_2 inventory. The fossil CO_2:CO emission ratios exhibit relatively little spatial variability, but emission rates can vary over small spatial scales (e.g., at urban/rural boundaries). Thus scaling by observed enhancements of CO gives a better estimate of the combustion signal. Photochemical loss of CO can be considered negligible over timescales of one day.

3.4. Biospheric Flux Model

[48] The biospheric model $\underline{\mathbf{F}}_{vegmod}$, following *Gerbig et al.* [2003b], was constructed with the aim of a simple representation (see equation (11)) that captures the diurnal variability in CO_2 fluxes. For each vegetation type v the CO_2 flux at (x_i, y_j, t_m) was modeled as the sum of a temperature (T) -dependent respiration term R_v and a photosynthetic uptake term GEE_v that is a function of the downward short-wave radiative flux (*SWRF*):

$$F_{vegmod,v}\left(x_i,y_j,t_m\right) = \lambda_{R,v}\cdot R_v\left(x_i,y_j,t_m\right) + \lambda_{GEE,v}\cdot GEE_i\left(x_i,y_j,t_m\right)$$

$$\text{where } R_v\left(x_i,y_j,t_m\right) = \alpha_v\left(x_i,y_j\right)\cdot\beta_v T\left(x_i,y_j,t_m\right) \quad (16)$$

$$GEE_v\left(x_i,y_j,t_m\right) = \alpha_v\left(x_i,y_j\right)\cdot\frac{a_v\cdot SWRF\left(x_i,y_j,t_m\right)}{b_v + SWRF\left(x_i,y_j,t_m\right)}$$

Table 2. Parameters Used for the Prior Biospheric Fluxes in Equation (16)

Experiment	Vegetation	Fitted Parameters	Eddy Covariance Sites Used for Fit; R^2
ND	forest	$\beta_i = 0.28$; $a_i = -50$; $b_i = 1864$	WLEF ($R^2 = 0.58$)
ND	cropland	$\beta_i = 0.26$; $a_i = -515$; $b_i = 9017$	Bondville ($R^2 = 0.80$)
WI	forest	$\beta_i = 0.28$; $a_i = -50$; $b_i = 1864$	WLEF ($R^2 = 0.58$)
WI	cropland	$\beta_i = 0.26$; $a_i = -515$; $b_i = 9017$	Bondville ($R^2 = 0.80$)
ME	forest	$\beta_i = 0.29$; $a_i = -41$; $b_i = 787$	Howland ($R^2 = 0.74$)
ME	cropland	$\beta_i = 0.26$; $a_i = -515$; $b_i = 9017$	Bondville ($R^2 = 0.80$)

$\alpha_v(x_i, y_j)$ is the fractional areal coverage at (x_i, y_j) for vegetation v (see below). We determined the parameters β_v, a_v, and b_v (Table 2) by fitting equation (16) to eddy covariance observations during July∼August 2000 at sites in the AmeriFlux network [*Baldocchi et al.*, 2001]. Surface T and *SWRF* were derived from EDAS assimilated fields. The simplicity of the biospheric model facilitates incorporation of information from eddy covariance observations and scaling of carbon fluxes to regional scales. Despite its apparent simplicity, the biospheric model accounted for at least 60% of the variance in hourly CO_2 fluxes at numerous AmeriFlux sites (Table 2).

[49] The areal coverage $\alpha_v(x_i, y_j)$ for each vegetation type was derived from the IGBP 1-km resolution vegetation data [*Belward et al.*, 1999], regridded to 1/6°Lat. × 1/4°Lon. resolution. We regrouped the 17 IGBP vegetation types into three dominant classes in the regions covered by the COBRA Lagrangian experiments: (1) forests (evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, deciduous broadleaf forest, and mixed forest) (2) croplands (croplands and cropland/natural vegetation mosaic), and (3) water (wetlands and water bodies).

[50] The scaling factors $\lambda_{R,v}$ and $\lambda_{GEE,v}$ for each vegetation type v were adjusted to minimize the deviation of the modeled biospheric signal $\Delta\widetilde{\underline{\mathbf{CO}}}_{2vegmod}$ from the observed signal using the Bayesian inverse method (equations (9), (12), and (14)). The elements of the state vector $\boldsymbol{\lambda}$ are composed of $\underline{\boldsymbol{\lambda}} = [\lambda_{GEE,forest}, \lambda_{R,forest}, \lambda_{GEE,crop}, \lambda_{R,crop}]^T$. The prior values ($\underline{\boldsymbol{\lambda}}_{prior}$) were set to 1.0.

[51] The net CO_2 flux for water bodies and wetlands was assumed to be zero. The upper limit of the magnitude in air-sea fluxes over the open ocean was estimated to be ∼0.1 μmoles/m^2/s based on recent pCO_2 data [*Lefevre et al.*, 1999]. The potential error for neglecting air-sea exchange fluxes will be included in the Bayesian inversion (see below). While carbon fluxes over the open ocean may not be applicable to fluxes from inland water bodies and wetlands, we will later show that this error is negligible in comparison with the other sources of errors because the associated areas covered by water are very small.

3.5. Specification of Error Covariance Matrices in Bayesian Inversion

[52] We calculated $\underline{\mathbf{S}}_{prior}$, the error covariance matrix for prior estimates of $\boldsymbol{\lambda}$, from comparison between modeled CO_2 fluxes and observed eddy covariance values at

AmeriFlux sites. The spatiotemporal correlations between different regions of the footprint were roughly accounted for by integrating the daytime and nighttime residuals over one day—the approximate timescale over which observed signals due to fluxes in the footprint are integrated—and dividing by the one-day integrated daytime and nighttime net ecosystem exchange to derive prior errors in $\lambda_{GEE,v}$ and $\lambda_{R,v}$, respectively [*Gerbig et al.*, 2003b]. Errors from spatial extrapolation were incorporated into $\underline{\underline{\mathbf{S}}}_{prior}$ for the forest class by imposing biospheric parameters derived at a single tower—WLEF for the WI experiments and Howland for the ME experiment—to other eddy covariance sites in different forests and calculating the residuals between modeled fluxes using imposed parameters and observed CO_2 fluxes. The other eddy covariance sites were: Blodgett, BOREAS NSA Black Spruce, BOREAS SSA Aspen, Duke, Harvard, Metolius, Niwot, Walker Branch, Willow Creek, and Wind River. The resulting prior uncertainties are 1.43 and 0.48 for $\lambda_{GEE,forest}$ and $\lambda_{R,forest}$ in WI and 0.66 and 0.46 for $\lambda_{GEE,forest}$ and $\lambda_{R,forest}$ in ME.

[53] We arbitrarily increased the uncertainty in scaling factors for the cropland class by a factor of 10, to 2.89 for $\lambda_{GEE,crop}$ and 3.74 for $\lambda_{R,crop}$, in order to reflect the fact that only a single cropland flux site, at Bondville, was available (Table 2) while the footprint sampled diverse crop types and management regimes.

[54] The measurement error covariance matrix $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ is treated as the sum of different components:

$$\underline{\underline{\mathbf{S}}}_{\varepsilon} = \underline{\underline{\mathbf{S}}}_{foss} + \underline{\underline{\mathbf{S}}}_{part} + \underline{\underline{\mathbf{S}}}_{eddy} + \underline{\underline{\mathbf{S}}}_{aggr} + \underline{\underline{\mathbf{S}}}_{water} + \underline{\underline{\mathbf{S}}}_{miss}. \quad (17)$$

Covariances reduce the degrees of freedom for the measurement errors and serve as prior information (or constraint) [*Brillouin*, 1956; *Rodgers*, 2000], but are currently not well known for the error sources in equation (17) at the regional scale. Thus we assumed no covariances in the errors making up $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ to minimize ad-hoc constraints on these errors. Note that $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ does not include contributions from instrument errors, which are random and independent, thus expected to largely cancel out, becoming negligible in comparison to the other error sources, when integrated in the calculation of column CO_2 amounts.

[55] $\underline{\underline{\mathbf{S}}}_{foss}$ is the error in determining fossil fuel contributions $\Delta\widetilde{CO_2}_{foss}$ (equation (15)). We assumed a large (conservative) standard deviation of 30% for the inventory-based ratio of fossil CO_2:CO enhancements $\delta\widetilde{CO_2}_{foss,grid}/\delta\widetilde{CO}_{grid}$; errors in $\langle F_{CO}\rangle$ primarily resulted from uncertainties in $\Delta\widetilde{CO}_{up}$ in equation (17) (see $\underline{\underline{\mathbf{S}}}_{miss}$ below).

[56] $\underline{\underline{\mathbf{S}}}_{part}$ is the error arising from using a finite number of random particles in STILT, ~13% for a typical signal in the mixed-layer for 100 particles [*Gerbig et al.*, 2003a]. $\underline{\underline{\mathbf{S}}}_{eddy}$ specifies the fluctuations in column-integrated CO_2 due to contributions from turbulent eddies, observed to be ~0.2 ppmv [*Gerbig et al.*, 2003a].

[57] $\underline{\underline{\mathbf{S}}}_{aggr}$ refers to the ''aggregation error'' arising from aggregating heterogeneous fluxes into a single flux [*Kaminski et al.*, 2001]. *Gerbig et al.* [2003b] demonstrated that a rough estimate of the aggregation error of CO_2 can be derived from the observed ''representation error''—i.e., the deviations between a point observation and a value averaged over a specific grid size [*Gerbig et al.*, 2003a]. From this result we can derive the corresponding aggregation error of 1 ppmv for fluxes at spatial scales of ~100 km for these experiments (see Figure 10 of *Gerbig et al.* [2003b]).

[58] $\underline{\underline{\mathbf{S}}}_{water}$ results from neglecting the carbon fluxes between the ocean and the atmosphere. We used the upper limit flux for oceanic fluxes of 0.1 μmole/m^2/s as the standard deviation of these uncertainties [*Lefevre et al.*, 1999] and combined these fluxes with the footprint contribution from water to arrive at the corresponding errors. We see later that $\underline{\underline{\mathbf{S}}}_{water}$ is much smaller than other error sources and remains negligible even if the error is increased by two orders of magnitude, because the footprints do not include large areas of open water.

[59] $\underline{\underline{\mathbf{S}}}_{miss}$ refers to the error due to the spatial mismatch between the location of upstream air parcels and the actual sampled locations that gives rise to uncertainties in the advected upstream tracer concentrations $CO_{2\ up}$. We made a conservative estimate of $\underline{\underline{\mathbf{S}}}_{miss}$. The spatial mismatch arose from shifts in winds between forecasted and analyzed meteorology as well as logistical limitations that prevented complete sampling of the particle locations. We first determined the mean distance $\overline{\Delta x}$ separating the sampled upstream cross-section from the particle locations during the same hour simulated with the assimilated winds. Then the mean observed gradient of $\widetilde{CO_2}_{up}$ in the upstream cross-section was calculated over this distance $\overline{\Delta x}$: $\left(\partial\widetilde{CO_2}_{up}/\partial x\right)\Big|_{\overline{\Delta x}}$. The absolute value of $\left(\partial\widetilde{CO_2}_{up}/\partial x\right)\Big|_{\overline{\Delta x}}$ was multiplied by the mean separation distance to derive the potential variability and uncertainty in $\widetilde{CO_2}_{up}$ that would exist over $\overline{\Delta x}$:

$$\delta\widetilde{CO_2}_{up} = abs\left(\frac{\partial\widetilde{CO_2}_{up}}{\partial x}\Big|_{\overline{\Delta x}}\right)\overline{\Delta x}. \quad (18)$$

The conservative nature of this error estimate arises from the fact that equation (18) assumes the errors to be correlated and additive; no cancellation of errors occurs from random fluctuations [*Taylor*, 1997]. The inherent assumption is that the tracer gradient observed in the upstream cross-section is similar to the unobserved gradient between the cross-section and the locations of the air parcels. The squared values of $\delta\widetilde{CO_2}_{up}$ were used as the diagonal elements of $\underline{\underline{\mathbf{S}}}_{miss}$. The mean separation distance between particles and the location of the upstream cross-section in COBRA was 50 km, smaller than the lengths of all the measured cross-sections, so we argue that the gradients in the unobserved parts probably do not require extensive extrapolation and are unlikely to be drastically different from the observed values.

4. Results

4.1. Results of STILT Simulation

4.1.1. Lagrangian "Matches" Between Upstream and Downstream

[60] We computed the locations of particles reaching the downstream receptors using STILT driven by EDAS meteorology. The distributions of particles at the times of the upstream flights are shown in Figures 4a–9a along with the flight tracks for both the downstream and upstream sampling times (see Table 1). These particles determine $\underline{\mathbf{I}}$, $\underline{\mathbf{f}}$, and the matching with $\underline{\mathbf{CO}}_{2_{i0}}$ providing estimates

for $\underline{\mathbf{CO}}_{2_{up}}$ (equation (10)), the advected upstream tracer concentrations.

[61] The effects of wind shear are evident in the particle distributions for WI#2, WI#3, and ME, particularly where particles originating at higher altitudes (e.g., 2100 m ASL) are disjoint from the rest of the particle ensemble. These particles traveled above the PBL, isolated from drag at the surface. Consequently, these particles were separated from the rest of the ensemble.

[62] In general the aircraft sampled only a subset of the upstream particle locations, as compared after the fact with the particle locations simulated using analyzed meteorology (Figures 4a–9a). The mean separation between the particle ensemble and the location of the sampled upstream cross-section was ~50 km. For example, particles reaching the receptor were sampled upstream on the southeastern part of the morning flights in North Dakota (Figure 4a, left panel). Much better overlap was found for particles started from higher altitudes, in the free troposphere, as verified by the presence of the same layer depleted in CO in both the upstream and downstream observations (Figure 4b).

[63] Mismatches were caused by wind shifts between the forecast and analysis: forecasted Eta winds in southern ND were northeasterly, with minimal wind shear between the PBL and the free troposphere, but the EDAS assimilation showed that PBL winds were more easterly. Forecasts for Maine (Figure 9a) agreed well with assimilated winds, but a navigation error caused upstream sampling to depart significantly from particles reaching the receptor.

[64] In WI#1 (Figure 5) the analyzed winds transported the particle ensemble to the north as it traveled backward in time, missing the sampled upstream cross-section (left-hand panel of Figure 5a). A flux could be estimated by subtracting concentrations in the left portion of the downstream cross-section (Figure 5b), which had depleted CO_2 and elevated CO concentrations, from the right portion of the upstream cross-section with elevated CO_2 and low CO concentrations. This procedure results in large CO_2 uptake (-40 μmole/m^2/s) and large CO emissions, almost an order of magnitude larger than the values in the CO emission grid. This result appears to be spurious, resulting from errors in the wind field, probably due to the inability of the EDAS to accurately model winds near the Great Lakes. A comparison of EDAS winds with the aircraft-observed winds showed that a large bias of 3 m/s in the north-south direction was present at the location of the downstream cross-section.

[65] The large tracer gradients in CO_2, CO, H_2O, and θ observed in both the upstream and downstream cross-sections of WI#1 suggested the same air mass was sampled, i.e., that the forecasted winds were better than the assimilated in this case. The correct matching between upstream and downstream requires proper alignment between the gradients. We accomplished this by shifting particle locations parallel to the upstream cross-section and forcing the mean particle location to match the mean position of flight paths in the upstream cross-section (right-hand panel of Figure 5a). The resulting shift in particle locations was 59 km, c1ose to the estimate of 43 km if the 3 m/s bias accumulated over the 4 hours (Table 1) when particles traveled between the downstream and the upstream cross-sections. The resulting CO flux was lowered to more reasonable values (not shown), and the CO_2 uptake was diminished to ~ -17 μmole/m^2/s. Owing to the ad hoc nature of the correction, we excluded WI#1 in the Bayesian inversion.

4.1.2. Footprints

[66] The footprint elements $f(\underline{\mathbf{x}}_r, t_r | x_i, y_j, t_m)$ comprising $\underline{\underline{\mathbf{f}}}$ are shown in the right panel of Figures 4a and 6a–9a. $\underline{\underline{\mathbf{f}}}$ is shown for all τ, the number of hours separating the times when the downstream and upstream tracer concentrations were altered by surface fluxes. No footprint is shown for WI#1 because of the large transport errors discussed in the previous section.

[67] For daytime experiments τ simply refers to the number of hours separating the upstream and downstream observations. In the case of diurnal experiments τ refers to hours elapsed since the time of cessation in vertical mixing during the previous day—not directly observed and subject to uncertainties—up until the time of the next day's afternoon observations. We estimated τ for the diurnal experiments by examining time series of simulated PBL heights at the particle locations and choosing the hours during the previous day when rapid collapse in PBL heights was observed, ranging from 19 to 23 hours between the different diurnal experiments (Table 1). We will attempt to bound the errors due to uncertainties in τ by conducting separate inversions for minimum and maximum values of τ.

[68] The footprint is the spatial region where the Lagrangian experiments have leverage to constrain surface fluxes. In the daytime ME experiment the footprint was restricted to regions in Maine. For the diurnal experiments ND and WI#2 the footprint extends much further upstream from the location of the upstream cross-section, as the residual layer observations in these cross-sections are influenced by surface fluxes from the previous day, when air parcels were further upstream. Stagnation during the diurnal experiments WI#3 and WI#4 translated into footprints restricted to northern Wisconsin.

4.2. Measured and Modeled Flux Signals

[69] We divide $\Delta\widetilde{\underline{\mathbf{CO}}}_{2veg}$, $\Delta\widetilde{\underline{\mathbf{CO}}}_{2vegmod}$, and $\Delta\widetilde{\underline{\mathbf{CO}}}_{2foss}$ by τ (see equation (5)) and display the corresponding fluxes $\langle\underline{\mathbf{F}}_{veg}\rangle$, $\langle\underline{\mathbf{F}}_{vegmod}\rangle$, and $\langle\underline{\mathbf{F}}_{foss}\rangle$ in Figures 4c and 4d, and Figures 6c and 6d through 9c and 9d. Here positive fluxes represent emissions to the atmosphere and negative fluxes uptake by the biosphere. We see that fossil combustion emissions of CO_2 are generally small in comparison to the sum of forest and cropland fluxes, as expected during the growing season for the nonurban regions covered by the Lagrangian experiment.

[70] The total biospheric CO_2 fluxes $\langle\underline{\mathbf{F}}_{veg}\rangle$ inferred from the data are shown in Figures 4d–9d as black lines. In the case of ND (Figure 4d) a west-east gradient of net diel uptake was obtained, increasing in magnitude from -1.6 to -3.8 μmole/m^2/s. The surface vegetation was dominated by wheat during early August. The gradient in CO_2 flux was correlated with the land-use gradient, with more nonagricultural grassland on the west, with smaller expected uptake (National Agricultural Statistics Service, U.S. Department of Agriculture, http://www.usda.gov/nass/). The a priori modeled CO_2 flux (dashed line, right panel) was dominated

by contributions from croplands with a gradient similar to the observed, but with uptake rates significantly greater compared to observed values.

[71] The diurnal experiments WI#3 and WI#4 sampled air arriving at the WLEF tall tower from the southwest, showing uptake values between −3.3 and −1.7 μmole/m^2/s. The observed flux was −1.25 μmole/m^2/s at WLEF during the entire growing season in 1997 [*Davis et al.*, 2003]. The corresponding CO_2 flux measured at WLEF (shown as green arrows in Figures 6d–8d) for the same period of time from 23 to 24 August was −1.4 μmole m^{-2} s^{-1}. The daytime flux in ME was ca. −15 μmole/m^2/s (Figure 9d). Eddy covariance observations at Howland Forest [*Hollinger et al.*, 1999], whose location is indicated by the red triangle in Figure 9a, showed a daytime flux of ca. −17 μmole/m^2/s for the same period, similar to the values estimated using the Lagrangian budget. Most of the uptake can be attributed to forests (Figure 9c).

4.3. Bayesian Inverse Analysis: Constraining the Biospheric Model and Regional-Scale Fluxes

[72] The footprint of Lagrangian aircraft experiments is, by design, several orders of magnitude larger (see right-hand panel of Figures 6a–8a) than that sampled by eddy covariance measurements (~1 km^2), so direct comparisons of fluxes require careful interpretation. Thus the aircraft observations and the AmeriFlux data are used within the context of the analysis framework presented in this study and in *Gerbig et al.* [2003b] as part of a Bayesian inverse analysis.

[73] $\widetilde{\Delta \mathbf{CO}}_{2veg}$, shown in Figures 4d and 6d–9d as $\langle \underline{\mathbf{F}}_{veg} \rangle$, was used within the Bayesian inverse method to optimize scaling parameters $\underline{\boldsymbol{\lambda}}$ within the simple biospheric flux model $\underline{\mathbf{F}}_{vegmod}$ (equation (16)). Separate optimizations for $\underline{\boldsymbol{\lambda}}$ were conducted for the ND, WI, and ME experiments in each case using the minimum and maximum τ (Table 1) as noted above. Results from the Bayesian inverse procedure are plotted in Figures 4d and 6d–9d as inferred CO_2 fluxes (blue dashed lines) assuming maximum τ in the individual Lagrangian experiments. In the same figures are dashed lines denoting $\langle \underline{\mathbf{F}}_{vegmod} \rangle$, the total prior modeled biospheric flux— sum of the forest (green) and cropland (orange) fluxes plotted in Figures 4c–9c. The optimized fluxes in Figures 4d–9d show improved agreement to the observed values, as expected.

[74] Figure 10 shows the prior (white) and optimized $\underline{\boldsymbol{\lambda}}$ for cases assuming maximum (black) and minimum (grey) values of τ, with their corresponding error bars (1-σ). The reduction in prior uncertainty is pronounced for GEE, $\lambda_{GEE,crop}$ in ND, $\lambda_{GEE,crop}$ and $\lambda_{GEE,forest}$ in WI, and $\lambda_{GEE,forest}$ in ME. Almost no reduction in uncertainties of $\lambda_{GEE,forest}$ in ND and smaller uncertainty reductions for $\lambda_{GEE,crop}$ in ME were observed, as forest and cropland contributions to the total biospheric uptake were minimal in ND (Figure 4c) and ME (Figure 9c), respectively. The reduction in uncertainty was generally small for R ($\lambda_{R,v}$) for both forest and cropland in all of the regions. The daytime flights, which seldom sample the nighttime CO_2 buildup, provide little constraint on respiration, so it is not surprising that the daytime experiment in ME provided leverage on $\lambda_{GEE,v}$ but not on $\lambda_{R,v}$. The diurnal experiments in ND and WI (which excluded WI#1 due to large uncertainties in winds) provided daily integrated constraints on uptake, a very useful number for carbon budgets. This information principally constrains $\lambda_{GEE,v}$ rather than $\lambda_{R,v}$ because, according to the AmeriFlux data, daily integrated photosynthetic fluxes are notably larger than respiration during the growing season, when COBRA was conducted.

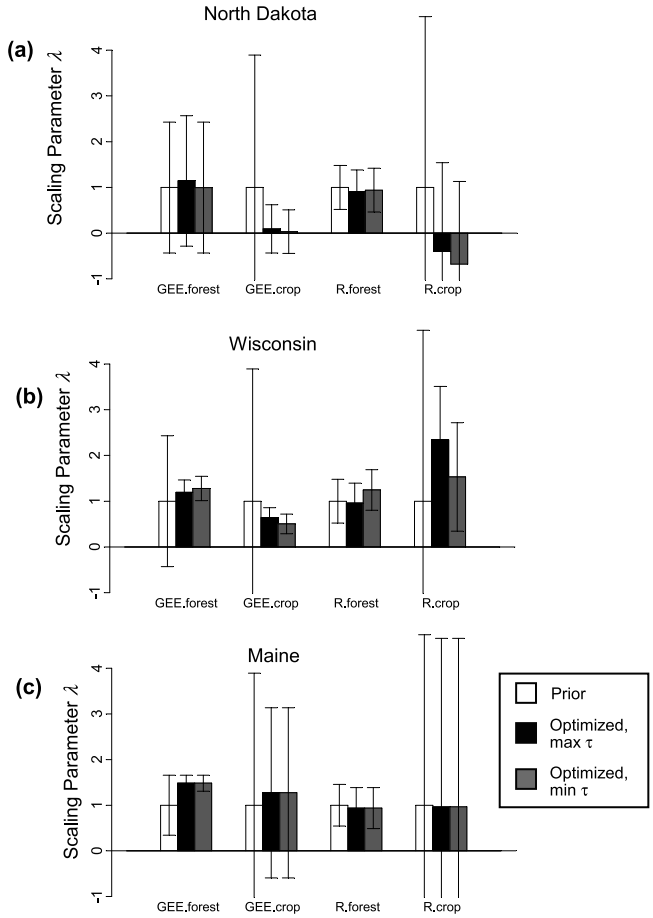

Figure 10. Results of the Bayesian inverse method solving for scaling parameters λ_v that adjust upward and downward the respiration (R_v) and photosynthetic (GEE_v) CO_2 fluxes (equation (16)) for vegetation type v. The prior λ_v are shown in white, while separate cases for optimized λ_v are derived for cases assuming maximum (black) and minimum (grey) values of τ. The error bars represent the 1-σ spread in λ_v.

[75] Optimized $\lambda_{GEE,crop}$ values are less than 1.0 in both ND and WI, suggesting the magnitude of the prior *GEE* was greater than actually observed. Overestimation of carbon uptake by croplands was found in all of the ND and WI experiments (Figures 4d–8d) due to several causes. Parameters for the cropland class were based solely on eddy covariance observations at the Bondville site, which measured corn in 2001 and soybeans in 2000, whereas wheat was the dominant crop in the ND experiment (National Agricultural Statistics Service, U.S. Department of Agriculture, http://www.usda.gov/nass/). Differences in characteristics of those crops certainly exist. Additionally, the wheat crop was close to being harvested (National Agricultural Statistics Service, U.S. Department of Agriculture, http://www.usda.gov/nass/) by the time of the ND Lagrangian

experiment and may not be photosynthesizing as quickly as earlier periods of rapid growth. Moreover, the "cropland/natural vegetation mosaic" class in the IGBP vegetation grid, widespread in the WI experiment, was treated as exclusively cropland cover in this analysis, which could lead to erroneous attribution of noncropland contributions as croplands. The photosynthetic carbon uptake from these vegetation types (especially wetlands) will likely not be as large as growing crops.

[76] In ME the optimized $\lambda_{GEE,forest}$ was ~1.5, suggesting that the prior *GEE* may have underestimated the actual value by 50%, albeit the optimized value lies within the error bars of the prior estimate. The EDAS-derived *SWRF* on this day was underestimated when compared to observations at Howland, highlighting uncertainties in meteorological variables other than wind vectors which propagate into the modeled carbon fluxes. If regional solar radiation were in fact higher than given by EDAS, our optimized values for $\lambda_{GEE,forest}$ would be too high.

[77] The differences in Figure 10 between optimized $\underline{\boldsymbol{\lambda}}$ using maximum and minimum values of τ were small for ND but large for WI, particularly for $\lambda_{R,crop}$. No differences were present in ME, as the ME flights were part of a daytime experiment with no uncertainty in τ. The sensitivity to τ in WI can be explained as follows: increasing τ resulted in hours with more solar radiation in the model, driving greater uptake and lowering the modeled $\Delta\widetilde{\underline{\mathbf{CO}}}_{2,veg}$ to more negative values. This was particularly pronounced for the WI#2 experiment, with a large cropland influence and sensitivity of *GEE* to solar radiation. To make the modeled $\Delta\widetilde{\underline{\mathbf{CO}}}_{2,veg}$ more positive and closer to the observed values, the inverse method greatly increased $\lambda_{R,crop}$, with its large prior errors, for maximum values of τ. $\lambda_{R,forest}$, in contrast, was more strongly constrained by the WLEF buildup of CO_2 on 24 August UT07, whose footprint covered more forest than croplands.

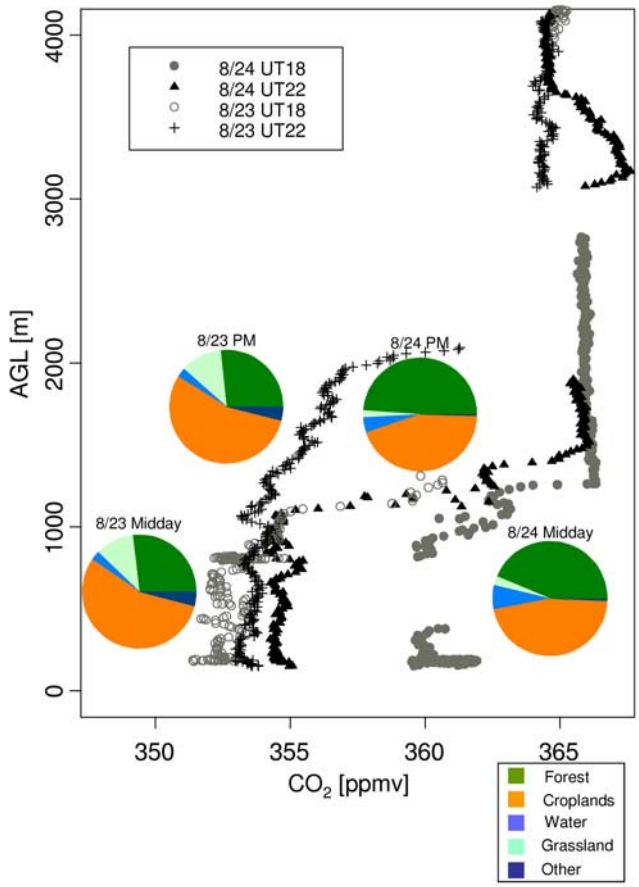

Figure 11. Vertical profiles conducted over the WLEF tall tower during 23 and 24 August and used to conduct one-dimensional Eulerian budgets for comparison with results from the Lagrangian method. The pie charts show footprint contributions from forest, croplands, water bodies, grasslands, and remaining IGBP classes as simulated by the STILT model and integrated over a 3-day travel period starting from WLEF.

5. Discussion

5.1. Comparison With One-Dimensional Eulerian Budget Method

[78] We compare the Lagrangian budget for the biospheric CO_2 flux, $\langle F_{veg} \rangle$, with values from a conventional one-dimensional Eulerian approach [*Denmead et al.*, 1996; *Kuck et al.*, 2000; *Levy et al.*, 1999; *Lloyd et al.*, 2001]. The main difference with the Lagrangian budget is that no upstream profile was used in the calculation; instead, changes in column amounts from vertical profiles observed over the same location as part of the WI experiments, at WLEF (Figure 11), were used to calculate fluxes from a one-dimensional budget.

[79] The profiles from the midday and afternoon on 24 and 23 August are shown in Figure 11. The data gaps resulted from in-flight calibrations by the CO_2 sensor and were filled by linear interpolation. The flux for a one-dimensional treatment, calculated from the change in column CO_2 between the midday and afternoon of 23 August was -11.8 μmole m^{-2} s^{-1}, lying within the errors of -17.5 μmole m^{-2} s^{-1}, the result from the Lagrangian budget. But on the following day, the increase in column CO_2 between the afternoon of 23 and 24 August implies net CO_2 release of 0.97 μmole m^{-2} s^{-1} over a diurnal period, opposite in sign and significantly different than the net flux of -2.3 μmole m^{-2} s^{-1}, suggesting uptake, derived from the WI#3 experiment.

[80] We used the STILT model to compute $\underline{\mathbf{f}}$ weighted by areal coverage from forest, croplands, water bodies, grasslands, and remaining classes, integrated over the 3-day travel period and displayed in Figure 10 as pie charts indicating percentages of the total footprint contribution from each vegetation type. The relative footprint contributions of different vegetation types were almost identical between the midday and afternoon profiles of 23 August, when the Lagrangian and one-dimensional methods yielded more similar results. In contrast, the afternoon profile from 24 August shows much larger contributions from forest, with a corresponding decrease in the contributions from croplands. The lower column amount of CO_2 on the afternoon of 23 August is likely related to higher uptake from croplands. This example highlights the need for caution when using a one-dimensional budget method, as differences in upstream vegetation contributions render

erroneous the homogeneity assumption necessary for the method [*Lin et al.*, 2003].

5.2. Potential for Constraining Carbon Fluxes

[81] The COBRA analyses have illustrated the potential but also the challenges and need for future improvements to the current application of the analysis framework to constrain carbon fluxes. The Lagrangian airborne experiments reduced prior errors in *GEE* (Figure 11) when incorporated into observational constraints within the context of a Bayesian inversion. The exact amount of uncertainty reduction, i.e., the retrieved information [*Rodgers*, 2000; *Shannon and Weaver*, 1963], clearly depends on the prior uncertainties. We attempted a conservative estimate (upper estimates of prior uncertainties) of prior errors, so the reduction in uncertainty may be overestimated. On the other hand, we also attempted to conservatively estimate the measurement errors $\underline{\underline{\mathbf{S}}}_{\varepsilon}$, particular for $\underline{\underline{\mathbf{S}}}_{miss}$ (error due to spatial mismatch between the location of upstream air parcels and actual sampled locations), so the actual degree of observational constraint to reduce uncertainties may be higher than applied in this study.

[82] The biospheric model $\underline{\mathbf{F}}_{vegmod}$, using the optimized scaling parameters $\underline{\boldsymbol{\lambda}}$, can be driven with environmental drivers *T* and *SWRF* to derive regional scale carbon fluxes with reduced uncertainty. The posterior errors in $\underline{\boldsymbol{\lambda}}$ (Figure 10) provide direct estimates of remaining uncertainties in these fluxes. In this way our receptor-oriented analysis framework made use of observations from Lagrangian experiments combined with ground-based eddy covariance measurements to provide estimates and the associated errors of regional scale carbon fluxes, estimates not currently available from alternative methods.

[83] The lack of constraints on respiration fluxes indicate the need for more complete sampling of the CO_2 buildup over the night in future experiments. One option is to have closer coordination between airborne observations and tall tower-based CO_2 measurements. The ground-based observations on the WLEF tall tower captured the nighttime buildup of CO_2 from respiration, poorly sampled by the aircraft due to difficulties associated with flights during the nighttime and within the early morning shallow PBL. Alternatively, ''missed approaches'' enable the aircraft to descend towards airports—into the shallow mixed layer—to measure the nighttime CO_2 accumulation, and then climb out again.

[84] The potential of Lagrangian experiments can be further realized in the future by reducing the size of the measurement errors (Table 3). The total measurement error $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ was dominated by contributions from uncertainties in the upstream $\underline{\underline{\mathbf{S}}}_{miss}$, responsible for over 50% of the total error. When converted into errors in the column averaged mixing ratio, $\underline{\underline{\mathbf{S}}}_{miss}$ resulted in mean uncertainties (square root of the diagonal elements) of over 1 ppmv. The prescribed uncertainty of 1 ppmv for the aggregation error $\underline{\underline{\mathbf{S}}}_{aggr}$ accounted for the second largest percentage and dominated the total error along with $\underline{\underline{\mathbf{S}}}_{miss}$. $\underline{\underline{\mathbf{S}}}_{water}$ was negligible in all experiment regions, at less than 0.001% of the total.

[85] $\underline{\underline{\mathbf{S}}}_{miss}$ is the largest source of measurement error throughout most of the Lagrangian experiments, because the strong spatial heterogeneity in CO_2 over the continent [*Gerbig et al.*, 2003a] translates into large uncertainties in $\widetilde{CO_2}_{up}$ (equation (18)) whenever the upstream sampling locations are separated from the locations of upstream air parcels (Figures 4a–9a). The mean separation distance between the particle ensemble and the location of the upstream cross-section was 50 km, out of which only 14 km was due to dispersion. This is the minimum mismatch resulting from the two-dimensional sampling pattern comprising the observed cross-sections. The remaining mismatch, over 2/3 of the total, was due to operational limitations encountered in the implementation of the flight plan as well as errors in predicting upstream air parcel locations. Improvements in the accuracy of flight planning for future Lagrangian experiments are required. The flight planning tool should incorporate advances in meteorological forecasting—e.g., the NCEP operational Eta 12 km model or the Weather Research and Forecasting (WRF) community model [*Michalakes et al.*, 2001]. The use of forecasted meteorology from multiple models to derive model-to-model spread as an indication of the forecast error helps to identify situations with unpredictable winds and enables planning of flight tracks that can span the model-to-model spread to reduce sensitivity to forecast errors. These improvements have been implemented as part of the recent 2003 COBRA campaign over North America and will be reported in a future publication.

[86] The analysis presented here represents a step forward in using data with high-frequency variability over the continent in order to constrain carbon fluxes, minimizing errors and loss of information associated with temporal aggregation [*Law et al.*, 2004; *Peylin et al.*, 2002]. However, errors remain from spatial aggregation of fluxes over large regions in order to reduce the degrees of freedom needed to be constrained [*Kaminski et al.*, 2001]. As mentioned above, the resulting aggregation error $\underline{\underline{\mathbf{S}}}_{aggr}$ is the second largest source of measurement errors (Table 3). One approach to reduce the contribution from $\underline{\underline{\mathbf{S}}}_{aggr}$ is to solve for fluxes at higher resolutions and to incorporate prior information through spatiotemporal covariances between the fluxes [*Peylin et al.*, 2001]. Future efforts to

Table 3. Errors in the Separate Terms Comprising the Measurement Error $\underline{\underline{\mathbf{S}}}_{\varepsilon}$ Used in the Bayesian Inverse Method (see Equation (17))[a]

Error Type	Experiment Region	Percent of Total Variance	Uncertainty in Column CO_2, ppmv
$\underline{\underline{\mathbf{S}}}_{miss}$	ND	51.01	1.06
$\underline{\underline{\mathbf{S}}}_{aggr}$	ND	39.23	1.00
$\underline{\underline{\mathbf{S}}}_{part}$	ND	6.63	0.41
$\underline{\underline{\mathbf{S}}}_{foss}$	ND	1.55	0.17
$\underline{\underline{\mathbf{S}}}_{eddy}$	ND	1.57	0.20
$\underline{\underline{\mathbf{S}}}_{miss}$	WI	68.05	1.42
$\underline{\underline{\mathbf{S}}}_{aggr}$	WI	25.00	1.00
$\underline{\underline{\mathbf{S}}}_{part}$	WI	4.23	0.41
$\underline{\underline{\mathbf{S}}}_{foss}$	WI	1.72	0.25
$\underline{\underline{\mathbf{S}}}_{eddy}$	WI	1.00	0.20
$\underline{\underline{\mathbf{S}}}_{miss}$	ME	69.99	1.37
$\underline{\underline{\mathbf{S}}}_{aggr}$	ME	24.81	1.00
$\underline{\underline{\mathbf{S}}}_{part}$	ME	4.19	0.41
$\underline{\underline{\mathbf{S}}}_{foss}$	ME	0.02	0.02
$\underline{\underline{\mathbf{S}}}_{eddy}$	ME	0.99	0.20

[a]The contribution from each error term is shown as both the percentage of the total variance (sum of all the diagonal elements in $\underline{\underline{\mathbf{S}}}_{\varepsilon}$) and the resulting uncertainty in the column-averaged CO_2 concentration. $\underline{\underline{\mathbf{S}}}_{water}$ was negligible in all cases, at less than 0.001% of the total.

reduce the aggregation error would also necessitate the use of a biospheric model that improves upon the crude simplification in this study, which divided the vegetation into only two classes—forest and croplands—and assumed that the biosphere behaved identically within each class. Additional information such as the crop planting and harvesting cycles, remote sensing datastreams such as MODIS-derived leaf area [*Myneni et al.*, 2002] and Enhanced Vegetation Index (EVI) [*Huete et al.*, 1997] or soil moisture from the HYDROS instrument (scheduled for launch in 2006) [*Reichle et al.*, 2001] would help improve the veracity of the biospheric representation. Direct optimization of biospheric parameters [*Kaminski and Heimann*, 2001] embedded within a sophisticated biospheric model, rather than solving for simple scaling parameters as presented in this study, is expected to offer a further step forward. The biospheric model provides powerful constraints on the inversion that are more natural than the ad hoc constraint generated by aggregation over crude vegetation classes.

[87] The Lagrangian method presented here employs intensive aircraft flights which are necessarily restricted in temporal coverage. Deriving carbon fluxes at longer timescales—up to a year—requires the use of the Lagrangian observations within the context of a large-scale research effort like the North American Carbon Program [*Wofsy and Harriss*, 2002], which seeks to bring together intensive experiments (such as Lagrangian aircraft experiments), long-term observational sites, and modeling efforts. The Lagrangian method provides a direct, independent constraint on fluxes that can be used to test fluxes derived from other methods—e.g., regional inverse estimates from CO_2 observed by a long-term observational network. The long-term observations can then be used to estimate fluxes from times when the Lagrangian observations are not available. The Lagrangian method can also be used to optimize the biospheric model at selected times of the year. The biospheric model could then be used as an "interpolator" that captures the temporal dynamics of carbon fluxes during other times. To accomplish this, the biospheric model would need to incorporate the aforementioned improvements to enhance its ability to capture the temporal dynamics of carbon fluxes.

[88] The issue of errors in the assimilated meteorological datasets remains unresolved in this study. EDAS may have underestimated radiation in the ME experiment, leading to underestimation in the modeled *GEE* (Figure 9d). Erroneous winds were clearly seen in the WI#1 experiment (Figure 5). To properly account for errors in the meteorological variables, direct comparisons with observations need to be carried out. For instance, direct comparisons of assimilated winds with radiosonde observations define error statistics that characterize the magnitude of the transport errors as well as how they correlate spatially and temporally. The error statistics could then be incorporated into the motion of STILT particles to propagate errors arising from incorrect transport (J. C. Lin et al., manuscript in preparation, 2004).

6. Summary and Conclusions

[89] We have outlined a receptor-oriented analysis framework to design and analyze Lagrangian experiments for quantifying regional scale fluxes of trace gases. STILT served as the natural tool for elucidating the locations of air parcels upstream from receptors in the PBL for flight planning and data analysis purposes, with its backward-time formulation and explicit treatment of turbulent transport to characterize effects from dispersion and wind shear. The observations of upstream and downstream concentrations provide direct measurements of regional fluxes ("Lagrangian budget"), enhancing the potential to provide tighter constraints on regional scale fluxes than from previous experiments—e.g., one-dimensional budgets.

[90] In this paper we illustrated the use of the framework for a case study, applying the framework to constrain regional scale carbon fluxes as part of the COBRA aircraft campaign. The constraints available from differences between upstream and downstream CO_2 observations from Lagrangian experiments provided estimates of regional scale carbon fluxes, with especially effective constraints on 24-hour mean CO_2 exchange and on large-scale rates for photosynthesis. Constraints on respiration were weak, but this defect could be overcome if more tall towers were available to measure the nocturnal buildup of CO_2 below aircraft operating altitudes. The framework incorporates information from eddy flux towers as priors in a Bayesian inverse analysis, thus using small-scale data in a consistent manner to help constrain large-scale fluxes. We identified current sources of uncertainties and outline clear steps to be undertaken to further realize the potential of this approach: improved flight planning, uncertainty analyses of meteorological variables, enhanced sophistication of the biospheric model, and incorporation of satellite data such as the Enhanced Vegetation Index.

[91] The analysis framework introduced in this study can be extended to atmospheric species other than CO_2. For instance, experiments can be planned so that sampling occurs upstream and downstream of a city to quantify urban pollutants, helping to minimize the advection component of the tracer budget and elucidate the surface emissions or chemical transformations. For pollutant species the surface flux model would not be a biospheric model, but a model of pollutant emissions and their chemical transformations, driven by variables such as temperature, population, and radiation. Such experiments have already been attempted for chemically active species like ozone, e.g., the TACIA study [*Kley et al.*, 1998], and for aerosols, the Aerosol Characterization Experiment (ACE) [*Bates et al.*, 1998]. The flight planning tool and footprint analysis introduced in this paper would provide unique new information, complementing the neutral balloons used in a study like ACE.

[92] **Acknowledgments.** We thank researchers in the AmeriFlux network for making available the eddy covariance observations. We gratefully acknowledge Ken Davis for comments on the manuscript and NOAA Air Resources Laboratory (ARL) for the provision of the EDAS meteorological files used in this publication. COBRA was supported by the following U.S. agencies: National Science Foundation [ATM-9821044], Department of Energy [DE-FG02-98ER62695], National Aeronautics and Space Administration [NAG5-7950], and National Oceanic and Atmospheric Administration [NA06GP0406]. J.C.L. was supported by the NASA Earth System Science Fellowship program.

References

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

Baldocchi, D. D. (2003), Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present, and future, *Global Change Biol.*, *9*, 479–492.

Baldocchi, D. D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*, 2415–2435.

Bates, T. S., B. J. Huebert, J. L. Gras, F. B. Griffiths, and P. A. Durkee (1998), International Global Atmospheric Chemistry (IGAC) project's First Aerosol Characterization Experiment (ACE 1): Overview, *J. Geophys. Res.*, *103*(D13), 16,297–16,318.

Belward, A. S., J. E. Estes, and K. D. Kline (1999), The IGBP-DIS global 1-km land-cover data set DISCover: A project overview, *Photogramm. Eng. Remote Sens.*, *65*(9), 1013–1020.

Benkovitz, C. M., M. T. Scholtz, J. Pacyna, L. Tarrason, J. Dignon, E. C. Voldner, P. A. Spiro, J. A. Logan, and T. E. Graedel (1996), Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.*, *101*(D22), 29,239–29,253.

Black, T. L. (1994), The new NMC mesoscale Eta-model—Description and forecast examples, *Weather Forecasting*, *9*(2), 265–278.

Boering, K. A., B. C. Daube, S. C. Wofsy, M. Loewenstein, J. R. Podolske, and E. R. Keim (1994), Tracer-tracer relationships and lower stratospheric dynamics: CO_2 and N_2O correlations during SPADE, *Geophys. Res. Lett.*, *21*(23), 2567–2570.

Brillouin, L. (1956), *Science and Information Theory*, 320 pp., Academic, San Diego, Calif.

Chameides, W. L., P. S. Kasibhatla, J. Yienger, and H. I. Levy (1994), Growth of continental-scale metro-agro-plexes, regional ozone production, and world food production, *Science*, *264*, 74–77.

Chou, W. W., S. C. Wofsy, R. C. Harriss, J. C. Lin, C. Gerbig, and G. W. Sachse (2002), Net fluxes of CO_2 in Amazonia derived from aircraft observations, *J. Geophys. Res.*, *107*(D22), 4614, doi:10.1029/2001JD001295.

Ciais, P., P. P. Tans, M. Trolier, J. W. C. White, and R. J. Francey (1995), A large northern hemisphere terrestrial CO_2 sink indicated by the $^{13}C/^{12}C$ ratio of atmospheric CO_2, *Science*, *269*, 1098–1102.

Cleugh, H. A., and C. S. B. Grimmond (2001), Modeling regional scale surface energy exchanges and CBL growth in a heterogeneous, urban-rural landscape, *Boundary Layer Meteorol.*, *98*, 1–31.

Conway, T. J., et al. (1994), Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network, *J. Geophys. Res.*, *99*(D11), 22,831–22,855.

Crutzen, P. J., and V. Ramanathan (2000), The ascent of atmospheric sciences, *Science*, *290*, 299–304.

Daube, B. C., K. A. Boering, A. E. Andrews, and S. C. Wofsy (2002), A high-precision fast-response airborne CO_2 analyzer for in situ sampling from the surface to the middle stratosphere, *J. Atmos. Oceanic Technol.*, *19*, 1532–1543.

Davis, K. J., P. S. Bakwin, C. Yi, B. W. Berger, C. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, *Global Change Biol.*, *9*, 1278–1293.

Denmead, O. T., M. R. Raupach, F. X. Dunin, H. A. Cleugh, and R. Leuning (1996), Boundary layer budgets for regional estimates of scalar fluxes, *Global Change Biol.*, *2*, 255–264.

Ehleringer, J. R., and C. B. Field (1993), *Scaling Physiological Processes: Leaf to Globe*, pp. 179–190, Academic, San Diego, Calif.

Environmental Protection Agency (EPA) (1993), Regional interim emission inventories (1987–1991), Research Triangle Park, N. C.

Fan, S., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi, and P. Tans (1998), A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models, *Science*, *282*, 442–446.

Gerbig, C., S. Schmitgen, D. Kley, A. Volz-Thomas, K. Dewey, and D. Haaks (1999), An improved fast-response vacuum-UV resonance fluorescence CO instrument, *J. Geophys. Res.*, *104*(D1), 1699–1704.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003a), Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, *108*(D24), 4756, doi:10.1029/2002JD003018.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003b), Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, *108*(D24), 4757, doi:10.1029/2003JD003770.

Gloor, M., S.-M. Fan, S. Pacala, J. Sarmiento, and M. Ramonet (1999), A model-based evaluation of inversions of atmospheric transport, using annual mean mixing ratios, as a tool to monitor fluxes of nonreactive trace substances like CO_2 on a continental scale, *J. Geophys. Res.*, *104*(D12), 14,245–14,260.

Goulden, M. L., J. W. Munger, S.-M. Fan, B. C. Daube, and S. C. Wofsy (1996), Exchange of carbon dioxide by a deciduous forest: Response to interannual climate variability, *Science*, *271*, 1576–1578.

Hollinger, D. Y., S. M. Goltz, E. A. Davidson, J. T. Lee, K. Tu, and H. T. Valentine (1999), Seasonal patterns and environmental control of carbon dioxide and water vapor exchange in an ecotonal boreal forest, *Global Change Biol.*, *5*, 891–902.

Huete, A. R., H. O. Liu, K. Batchily, and W. van Leeuwen (1997), A comparison of vegetation indices global set of TM images for EOS-MODIS, *Remote Sens. Environ.*, *59*(3), 440–451.

Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001: The Scientific Basis*, 944 pp., Cambridge Univ. Press, New York.

Kaminski, T., and M. Heimann (2001), Inverse modeling of atmospheric carbon dioxide fluxes, *Science*, *294*, 259–260.

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

Kley, D., C. Gerbig, and S. Schmitgen (1998), PI-report, in *Testing Atmospheric Chemistry in Anticyclones (TACIA)*, edited by O. Hov, final report, contract ENV4-CT95-0038, pp. 128–132, Res. and Dev. Programme Environ. and Clim. Forsch. Jülich, Jülich, Germany.

Kuck, L. R., et al. (2000), Measurements of landscape-scale fluxes of carbon dioxide in the Peruvian Amazon by vertical profiling through the atmospheric boundary layer, *J. Geophys. Res.*, *105*(D17), 22,137–22,146.

Law, R. M., et al. (1996), Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, *Global Biogeochem. Cycles*, *10*(4), 783–796.

Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting (2004), Data and modeling requirements for CO_2 inversions using high frequency data, *Tellus*, in press.

Lefevre, N., A. J. Watson, D. J. Cooper, R. F. Weiss, T. Takahashi, and S. C. Sutherland (1999), Assessing the seasonality of the oceanic sink for CO_2 in the northern hemisphere, *Global Biogeochem. Cycles*, *13*(2), 273–286.

Levin, S. A. (1992), The problem of pattern and scale in ecology, *Ecology*, *73*, 1943–1967.

Levy, P. E., A. Grelle, A. Lindroth, M. Molder, P. G. Jarvis, B. Kruijt, and J. B. Moncrieff (1999), Regional-scale CO_2 fluxes over central Sweden by a boundary layer budget method, *Agric. For. Meteorol.*, *98–99*, 169–180.

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.

Lloyd, J., et al. (2001), Vertical profiles, boundary layer budgets, and regional flux estimates for CO_2 and its $^{13}C/^{12}C$ ratio and for water vapor above a forest/bog mosaic in central Siberia, *Global Biogeochem. Cycles*, *15*(2), 267–284.

Marland, G., T. A. Boden, A. L. Brenkert, R. J. Andres, and J. G. J. Olivier (1997), CO_2 from fossil fuel burning: Updates on the magnitude, distribution, and uncertainty of emission estimates, paper presented at Fifth International Carbon Dioxide Conference, Queensland, Australia.

Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock (2001), Development of a next generation regional weather research and forecast model, in *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*, edited by W. Zwieflhofer and N. Kreitz, pp. 269–276, World Sci., River Edge, N. J.

Myneni, R. B., et al. (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, *83*, 214–231.

Newson, M. D., and I. R. Calder (1989), Forests and water resources: Problems of prediction on a regional scale, *Philos. Trans. R. Soc. London, Ser. B*, *324*, 283–298.

Peylin, P., P. Bousquet, and P. Ciais (2001), Response to "Inverse modeling of atmospheric carbon dioxide fluxes," *Science*, *294*, 259–260.

Peylin, P., D. Baker, J. Sarmiento, P. Ciais, and P. Bousquet (2002), Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO_2 data, *J. Geophys. Res.*, *107*(D19), 4385, doi:10.1029/2001JD000857.

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

Reichle, R., D. B. McLaughlin, and D. Entekhabi (2001), Variational data assimilation of microwave radiobrightness observations for land surface hydrologic applications, *IEEE Trans. Geosci. Remote Sens.*, *39*(8), 1708–1718.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, 238 pp., World Sci., River Edge, N. J.

Rogers, E., D. G. Deaven, and G. J. Dimego (1995), The regional analysis system for the operational early Eta-model: Original 80-km configuration and recent changes, *Weather Forecasting*, *10*(4), 810–825.

Running, S. W., and E. R. J. Hunt (1993), Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models, in *Scaling Physiological Processes: Leaf to Globe*, edited by J. R. Ehleringer and C. B. Field, pp. 141–158, Academic, San Diego, Calif.

Sarmiento, J. L., and S. C. Wofsy (1999), A U.S. carbon cycle science plan, Carbon and Clim. Working Group, U.S. Global Change Res. Program, Washington, D. C.

Schimel, D., et al. (2000), Contribution of increasing CO_2 and climate to carbon storage by ecosystems in the United States, *Science*, *287*, 2004–2006.

Schimel, D. S., et al. (2001), Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems, *Nature*, *414*, 169–172.

Schmitgen, S., et al. (2004), Carbon dioxide uptake of a forested region in southwest France derived from airborne CO_2 and CO measurements in a quasi- Lagrangian experiment, *J. Geophys. Res.*, *109*, D14302, doi: 10.1029/2003JD004335.

Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua (1996), A revised land surface parameterization (SIB2) for atmospheric GCMs. I. Model formulation, *J. Clim.*, *9*(4), 676–705.

Shannon, C., and W. Weaver (1963), *The Mathematical Theory of Communication*, 125 pp., Univ. of Ill. Press, Urbana.

Stephens, B. B., S. C. Wofsy, R. F. Keeling, P. P. Tans, and M. J. Potosnak (2000), The CO_2 budget and rectification airborne study: Strategies for measuring rectifiers and regional fluxes, in *Inverse Methods in Global Biogeochemical Cycles*, edited by P. Kasibhatla et al., pp. 311–324, AGU, Washington, D. C.

Tans, P. P., I. Y. Fung, and T. Takashi (1990), Observational constraints on the global atmospheric CO_2 budget, *Science*, *247*, 1431–1438.

Tans, P. P., P. S. Bakwin, and D. W. Guenther (1996), A feasible Global Cycle Observing System: A plan to decipher today's carbon cycle based on observations, *Global Change Biol.*, *2*, 309–318.

Taylor, J. R. (1997), *An Introduction to Error Analysis*, Univ. Science Books, Sausalito, Calif.

Wofsy, S. C., and R. C. Harriss (2002), The North American Carbon Program (NACP), NACP Comm. of the U.S. Interagency Carbon Cycle Sci. Program, U.S. Global Change Res. Program, Washington, D. C.

Wofsy, S. C., R. C. Harriss, and W. A. Kaplan (1988), Carbon dioxide in the atmosphere over the Amazon basin, *J. Geophys. Res.*, *93*(D2), 1377–1387.

Wofsy, S. C., M. L. Goulden, J. W. Munger, S.-M. Fan, P. S. Bakwin, B. C. Daube, S. L. Bassow, and F. A. Bazzaz (1993), Net exchange of CO_2 in a midlatitude forest, *Science*, *260*, 1314–1317.

A. E. Andrews and P. S. Bakwin, Climate Monitoring and Diagnostics Laboratory, NOAA, Boulder, CO 80303, USA.

B. C. Daube, C. Gerbig, J. C. Lin, and S. C. Wofsy, Deparment of Earth and Planetary Sciences and Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. (jcl@io.harvard.edu)

C. A. Grainger, Department of Atmospheric Sciences, University of North Dakota, Grand Forks, ND 58202, USA.

D. Y. Hollinger, Forest Service, U.S. Department of Agriculture, Northeast Research Station, Durham, NH 03824, USA.

B. B. Stephens, Atmospheric Technology Division, National Center for Atmospheric Research, Boulder, CO 80305, USA.