Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent:
2. Analysis of COBRA data using a receptor-oriented framework

C. Gerbig,[1] J. C. Lin,[1] S. C. Wofsy,[1] B. C. Daube[1,2] A. E. Andrews,[1,3] B. B. Stephens,[4] P. S. Bakwin,[5] and C. A. Grainger[6]

[1]   We present an analysis framework and illustrate its potential to constrain terrestrial carbon fluxes at the regional scale using observations of CO_2 and CO over North America acquired during the CO_2 Budget and Rectification Airborne (COBRA) study in 2000. The COBRA data set, presented in detail in a companion paper [*Gerbig et al.*, 2003] provides dense spatial coverage and extensive profiling in the lower atmosphere, revealing strong CO_2 signatures of land surface fluxes in the active and relic mixed layers of the atmosphere. We introduce a "receptor-oriented" analysis framework designed to quantitatively interpret the atmospheric signatures of surface processes by linking concentrations at measurement locations (receptors) to surface fluxes in upwind regions. The framework incorporates three main components: (1) the Stochastic Time-Inverted Lagrangian Transport (STILT) model, driven with assimilated winds and running backward in time to map out the source-receptor relationship (footprint) at high temporal and spatial resolution; (2) an observation-based lateral boundary condition for CO_2, resolving vertical and meridional gradients; and (3) a simple parameterization for biosphere-atmosphere fluxes that uses eddy covariance observations from the AmeriFlux network as prior estimates for fluxes. This framework allows quantitative comparison between the top-down constraint on fluxes from airborne observations of CO_2 with the bottom-up constraint of eddy flux measurements in a Bayesian synthesis inversion. The model is used to investigate the observed representation error (mismatch between point measurements and grid-cell-averaged values in models), evaluated in the companion paper, showing that unresolved spatial variability of surface fluxes gives rise to most of the representation error over the continent. Thus the representation error reflects the effect of aggregation errors. Discrepancies between simulated and observed CO_2 distributions are assessed to indicate where improvements are needed, including improved empirical representation of biosphere-atmosphere exchange process and better simulation of convective processes in atmospheric transport models.   *INDEX TERMS:* 0315 Atmospheric Composition and Structure: Biosphere/atmosphere interactions; 0322 Atmospheric Composition and Structure: Constituent sources and sinks; 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; 1610 Global Change: Atmosphere (0315, 0325); 1615 Global Change: Biogeochemical processes (4805); *KEYWORDS:* adjoint model, Lagrangian model, regional fluxes of CO_2

1.   Introduction

[2]   Continental CO_2 fluxes exhibit strong spatial and temporal variations: biospheric fluxes reverse sign between night (respiration) and day (respiration + photosynthetic uptake), and vary across the landscape with vegetation type and conditions. The atmosphere transports and modifies the signals from the surface fluxes, giving rise to a CO_2

---

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

[2]Now at Climate Monitoring and Diagnostic Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA

[3]Now at NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

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

[5]Climate Monitoring and Diagnostic Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA.

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

distribution which varies significantly on relatively small spatial and temporal scales. This variance leads to potentially large representation errors, uncertainties due to the comparison of point observations with grid-cell averages computed by models. The proximity of measurements to strong surface fluxes requires an analysis framework that resolves sources and sinks in the near field of the observations [*Lin et al.*, 2003], even if large-scale fluxes are the objective; otherwise the flux estimates may be invalid because of aggregation errors, the accounted extra weight of fluxes in the near field [*Kaminski et al.*, 2001]. The covariance between atmospheric mixing and fluxes on diurnal timescales ("diurnal rectifier effect" [*Denning et al.*, 1996]) needs to be accurately represented.

[3] *Gerbig et al.* [2003] investigated observations collected during the CO_2 Budget and Rectification Airborne (COBRA) study to characterize the spatial variability of atmospheric CO_2. This companion paper found that, in order to effectively utilize the signals from fluxes over the continent, a modeling framework with high spatial resolution (<100 km) was required, with full utilization of atmospheric concentration variations needing resolution of $\sim$30 km.

[4] The high-frequency signals of CO_2 over the continent, such as caused by diurnally changing fluxes or by variations in flow patterns together with spatial gradients in upstream fluxes, represent a rich source of information on source/sink distributions, when inversion techniques involving a transport model are applied [*Law et al.*, 2002]. A recent analysis by *Wang* [2003] applied Bayesian synthesis inversion to hourly CO_2 measurements made at Cape Grim, off the coast of the Australian Continent, to infer biospheric exchange fluxes. A significant reduction in uncertainty of fluxes was found for regions in southern Australia. Since the measurement location was $\sim$200 km from the continent, the data at Cape Grim are not directly affected by the strong spatial variability of upstream fluxes in the near-field, but only a limited portion of the continent could be assessed.

[5] This paper introduces an analysis framework designed to start addressing the problems encountered when using data collected over the continent, including representation and aggregation errors and the diurnal rectifier. The framework is intended to be a first step toward the "model-data fusion" envisioned in the North American Carbon Program (NACP) [*Wofsy and Harriss*, 2002] to assimilate the greatly expanded data set of CO_2 observations over the continent expected over the next few years.

[6] Figure 1 presents a schematic representation of the framework. It consists of three main components: a receptor-oriented transport model using analyzed meteorological field, a lateral tracer boundary condition using remote marine measurements and a Green's function for vertical propagation, and a simple biospheric flux model using AmeriFlux tower data [*Baldocchi et al.*, 2001] to provide high spatial and temporal resolution. The framework takes input from multiple data sources and provides as output optimized biospheric parameters and, in conjunction with meteorological fields, associated regional fluxes. We present a first application of the framework to COBRA CO_2 measurements.

[7] The high spatial and temporal resolution provides the potential to reduce representation and aggregation errors, and the explicit simulation of diurnally varying biospheric fluxes and turbulent mixing enables the diurnal rectifier to be represented. The application of the data analysis framework to the observations is done in several steps (Figure 1): (1) Influence functions are calculated for the measurement locations at high spatial and temporal resolution using the receptor-oriented Stochastic Time-Inverted Lagrangian Transport (STILT) model [*Lin et al.*, 2003]; these influence functions are equivalent to the adjoint of the transport model, in that they represent sensitivities of atmospheric concentrations to upstream surface fluxes or lateral boundary values. (2) The influence functions are coupled to emission inventories for fossil fuels and to background fields for CO_2 and CO, to derive the combustion CO_2 signal (involving measured CO enhancements above the background as a combustion tracer and inventory based CO_2/CO emission ratios) as well as the advected background CO_2. (3) The measurement-based CO_2 vegetation signal is calculated as the difference between measured CO_2 and the sum of advected CO_2 and combustion CO_2 signal. (4) Biospheric fluxes are modeled as responses to temperature and radiation from assimilated meteorological data; the responses to these meteorological drivers are keyed to data from eddy covariance flux towers in the AmeriFlux network [*Baldocchi et al.*, 2001] (5) Modeled CO_2 vegetation signals are derived by coupling the influence functions to the biosphere model. (6) Parameters of the biospheric flux model for key vegetation types are optimized in a Bayesian synthesis inversion to obtain a match between modeled and measured CO_2 vegetation signals. (7) As an illustration of the potential of the framework, estimates of regional fluxes for $\sim$1 week are derived by driving the optimized biosphere flux model with observed meteorological conditions.

[8] This paper is organized as follows: We start in section 2 with a description of the major elements of the data analysis framework designed to cope with, and learn from, the observed variability of atmospheric concentrations, including the receptor-oriented transport model, parameterizations for biosphere and fossil fuel fluxes, and a statistical model used to create the lateral boundary condition for CO_2 and CO from ground-based and airborne measurements to the west of the North American continent. Investigation of sensitivities of the transport model to key parameters is presented in section 3. The results of applying the receptor-oriented framework to COBRA data are discussed in section 4, with a presentation of the surface influence functions ("footprints") for COBRA (section 4.1), the derivation of the measurement-based CO_2 vegetation signal (section 4.2), a Bayesian synthesis inversion to derive scaling factors for the biosphere model (section 4.3), and an investigation of the factors regulating representation error [*Gerbig et al.*, 2003] using the receptor-oriented framework (section 4.3). In section 5 these results are discussed, with emphasis on discrepancies between the observed and modeled vegetation signals, and how these discrepancies relate to specific shortcomings of currently available meteorological data and other elements of the framework.

2. Description of the Analysis Framework

[9] An analysis framework is required that can adequately resolve the spatial variability of CO_2 as discussed in the

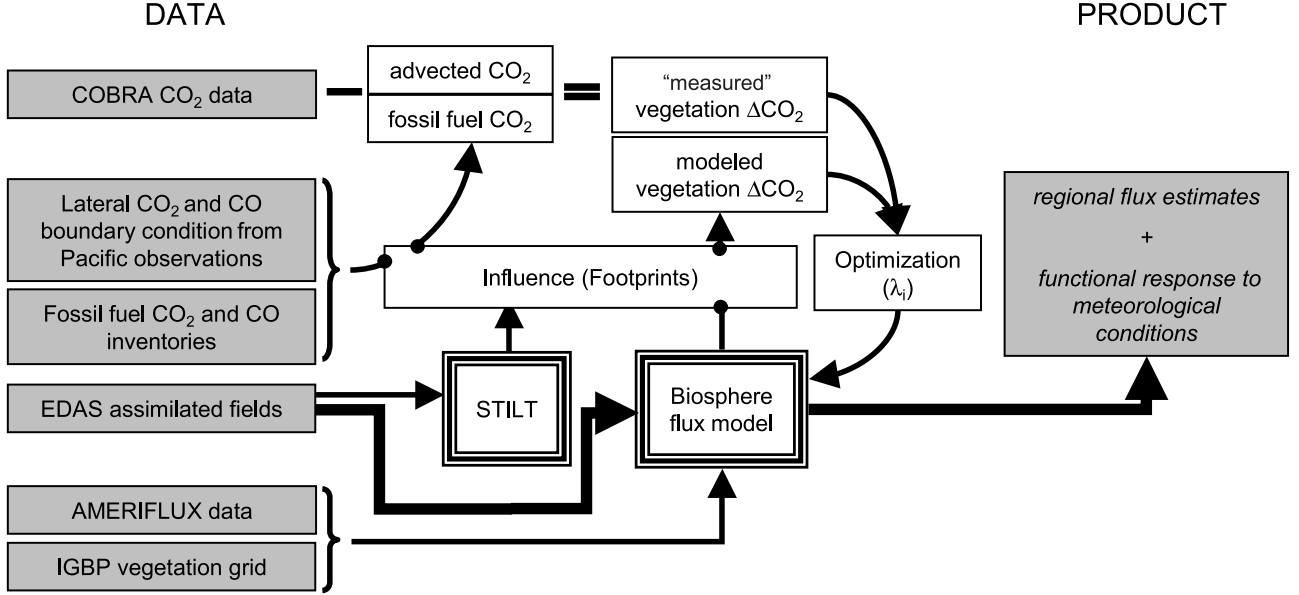DATA   PRODUCT

Figure 1. Schematic diagram of the data-analysis framework.

companion paper [*Gerbig et al.*, 2003], with minimal representation error, and extract the signal of terrestrial biospheric fluxes present in the aircraft observations (see Figure 3 in the companion paper). Further requirements of the analysis framework (Figure 1) are: (1) high temporal resolution to resolve diurnal cycles in terrestrial fluxes and boundary-layer dynamics, as well as taking into account synoptic events that introduce variations in atmospheric CO_2; (2) coupling to a boundary condition that connects tracer mixing ratios over the continent with global distributions, crucial for long-lived tracers like CO_2; (3) the ability to incorporate surface fluxes from the biosphere and fossil fuel combustion at high temporal and spatial resolution, providing the potential to minimize aggregation errors.

[10] The following sections describe the different elements of the analysis framework (Figure 1). We first introduce the receptor-oriented transport model (section 2.1) used to derive influence functions (footprints), which link the observations with upstream boundary conditions and surface fluxes. We then describe the lateral boundary conditions for CO_2 and CO mixing ratios (section 2.2), fossil fuel fluxes (section 2.3), and the biospheric flux model (section 2.4).

2.1. Receptor-Oriented Modeling Framework

[11] We developed a receptor-oriented atmospheric modeling framework (in the following referred to as ROAM) that quantitatively relates local measurements to the flux distribution upwind at high spatial and temporal resolution.

2.1.1. Theory

[12] The ROAM problem may be stated as follows: given a tracer concentration $C(\mathbf{x}_r, t_r)$ at location $\mathbf{x}_r$ measured at time t_r, what is the influence of a given surface flux element upstream? We introduce the influence functions $I(\mathbf{x}_r, t_r \mid \mathbf{x}, t)$, which quantitatively link surface sources or sinks $S(\mathbf{x}, t)$, for a conserved tracer emitted at location $\mathbf{x}$ and at time t, to the

tracer concentration at a receptor [*Holzer and Hall*, 2000; *Lin et al.*, 2003]:

$$C(\mathbf{x}_r, t_r) = \int_{t_0}^{t_r} dt \int_V d^3\mathbf{x} I(\mathbf{x}_r, t_r | \mathbf{x}, t) S(\mathbf{x}, t)$$
$$+ \int_V d^3\mathbf{x} I(\mathbf{x}_r, t_r | \mathbf{x}, t_0) C(\mathbf{x}, t_0) \qquad (1)$$

[13] The first term on the RHS represents changes in the concentration at the receptor due to surface fluxes in the domain V between initialization time t_0 and t_r. Influence I has units of inverse volume (a density) and, since we use mixing ratios for C (ppm), S has units of ppm/s. The second term is the contribution from advection of the initial tracer field $C(\mathbf{x}, t_0)$. Surface fluxes in equation 1 can be represented as interior (volume) sources or sinks [*Holzer and Hall*, 2000], placed between surface and a column height h, chosen to be smaller than the mixed-layer height z_i:

$$S(\mathbf{x}, t) = \begin{cases} \dfrac{F(x, y, t) \cdot m_{air}}{h \cdot \overline{\rho}(x, y, t)} & \text{for } z \leq h \\ 0 & \text{for } z > h \end{cases} \qquad (2)$$

[14] Here F is the surface flux in µmol/m²/s, $\overline{\rho}$ is column-averaged density of air, and m_{air} is the molar mass of air. To implement surface fluxes at finite temporal (Δt) and spatial resolutions (Δx, Δy), we integrate the first term of the RHS of equation (1) over the discrete temporal and spatial intervals:

$$\Delta C_{i,j,k}(\mathbf{x}_r, t_r) = F(x_j, y_k, t_i) \left[\frac{m_{air}}{h \cdot \overline{\rho}(x_j, y_k, t_i)} \right.$$
$$\left. \cdot \int_{t_i}^{t_i+\Delta t} dt \int_{x_j}^{x_j+\Delta x} dx \int_{y_k}^{y_k+\Delta y} dy \int_0^h dz I(\mathbf{x}_r, t_r | \mathbf{x}, t) \right] \qquad (3)$$

[15] $\Delta C_{i,j,k}(\mathbf{x}_r, t_r)$ is the change in tracer amount due to fluxes $F(x_j, y_k, t_i)$ emitted from a surface grid element j, k, and in the time interval between t_i and $t_i + \Delta t$. The column-averaged atmospheric mass density is assumed to be constant over the grid element. The term in brackets defines source-receptor elements that link surface fluxes to concentration changes at the receptor, denoted as the "footprint element" $f(\mathbf{x}_r, t_r| x_j, y_k, t_i)$. We refer to the spatial distribution of footprint elements for a given time interval as the "footprint".

2.1.2. STILT Model

[16] We extract footprint information from assimilated meteorological data using the newly developed Stochastic Time-Inverted Lagrangian Transport (STILT) model [Lin et al., 2003]. STILT simulates transport backward in time by an ensemble of representative particles released at a receptor. The particles represent air parcels of equal mass, transported by mean winds and subgrid turbulent winds calculated from surface sensible heat and momentum fluxes, extracted from the assimilated meteorological fields. The local density of these particles is directly related to the influence density. Given N_{tot} particles released from a receptor at $\mathbf{x}_r$ at time t_r, giving rise to particle density at location $\mathbf{x}$ and time t, $\rho(\mathbf{x}_r, t_r| \mathbf{x}, t)$, the influence I is given by $I(\mathbf{x}_r, t_r| \mathbf{x}, t) = \rho(\mathbf{x}_r, t_r| \mathbf{x}, t)/N_{tot}$. If we integrate I over a time interval Δt and volume element above surface grid cell (j, k), we find:

$$\int_{t_i}^{t_i+\Delta t} dt \int_{x_j}^{x_j+\Delta x} dx \int_{y_k}^{y_k+\Delta y} dy \int_0^h dz I(\mathbf{x}_r, t_r|\mathbf{x}, t) = \frac{1}{N_{tot}} \sum_{p=1}^{N_{tot}} \Delta t_{p,i,j,k} \quad (4)$$

[17] The RHS is the fraction of particles found in the volume $h \cdot \Delta x \cdot \Delta y$, and $\Delta t_{p,i,k,j}$ is the time which each individual trajectory p spends in this volume above cell j, k, during the time interval Δt. STILT employs operator splitting between horizontal advection (long time step, Δt), and vertical turbulence (short time step) (for details, see Lin et al. [2003]), hence the times $\Delta t_{p,i,k,j}$ are multiples of the fast timestep. The footprint elements $f(\mathbf{x}_r, t_r| x_j, y_k, t_i)$ in equation (3) are therefore given by

$$f(\mathbf{x}_r, t_r|x_j, y_k, t_i) = \frac{m_{air}}{h \cdot \overline{\rho}(x_j, y_k, t_i)} \frac{1}{N_{tot}} \sum_{p=1}^{N_{tot}} \Delta t_{p,i,j,k} \quad (5)$$

and the tracer signal due to upstream surface fluxes F ($C_{surface}$) is obtained by summing equation (3) over all backward timesteps (i) and surface grid elements (j, k):

$$
\begin{aligned}
C_{surface}(\mathbf{x}_r, t_r) &= \sum_{i,j,k} \Delta C_{i,j,k}(\mathbf{x}_r, t_r) \\
&= \sum_{i,j,k} f(\mathbf{x}_r, t_r|x_j, y_k, t_i) \cdot F(x_j, y_k, t_i) \\
&= \sum_{i,j,k} \frac{m_{air}}{h \cdot \overline{\rho}(x_j, y_k, t_i)} \cdot \frac{1}{N_{tot}} \left(\sum_{p=1}^{N_{tot}} \Delta t_{p,i,j,k} \right) \\
&\quad \cdot F(x_j, y_k, t_i)
\end{aligned}
\quad (6)
$$

[18] The footprint is similar to the adjoint of a Eulerian transport model [Errico, 1997]: each footprint element is equivalent to the sensitivity of the mixing ratio at a given receptor location with respect to a change in boundary (surface) flux. The approach of using a Lagrangian particle dispersion model to derive footprint information has multiple advantages: (1) Interpolation of winds down to the exact location of a measurement enables footprints to be derived at much higher spatial resolution than the driving meteorological data. This is important for the near field, close to the receptor, where influence from heterogeneous surface fluxes is strongest and footprint areas are small, providing the potential to minimize aggregation error. (2) The representation of influence with particle distributions avoids representation errors, since the volume represented by the particles at the time of the measurement is infinitesimal. However, a new type of "sampling" error is introduced because of the finite number of particles used to represent transport to an individual measurement location. (3) Modeling turbulent transport as the ensemble of stochastically transported particles more closely approximates the stochastic nature of air parcels transported by turbulence than typical parameterizations (e.g., diffusion coefficients). (4) The fact that the model runs backward in time makes it very efficient, such that only a single reversed-time model run is required to extract the spatially and temporally resolved footprint for emissions at all previous times.

[19] For analysis of the COBRA large-scale survey data the STILT model was driven by assimilated meteorological data from EDAS (ETA Data Assimilation System) [Rogers et al., 1995] which covers the United States and parts of Canada at 80 km horizontal resolution, 250 m vertical resolution close to the surface, and updates at intervals of 3 hours, supplemented by GDAS (Global Data Assimilation System) data covering the Northern Hemisphere at ~180 km resolution with 6-hourly updates. Each run was started using EDAS winds, and meteorological grids were changed as soon as one particle left the EDAS area. A receptor point was defined along the flight track of the two large-scale transects whenever the aircraft moved vertically by about 30 mbar or horizontally by 30 km. An ensemble of N_{tot} particles was released at each receptor point, with the value of N_{tot} determined from a sensitivity test (see section 3.1). Particle positions were calculated using appropriate time steps for advection (Courant number <0.25, between 3 and 120 minutes), and for turbulence (less than 10% of the Lagrangian timescale for turbulence, 0.1 seconds to 10's of seconds). Calculations for a particle were carried out until it crossed the western model boundary (145°W, see Figure 2) or after a maximum duration of 15 days.

[20] The particle trajectories were mapped onto a surface domain ("ROAM domain") covering the continental United States, Mexico and most of Canada (Figure 2). Most particles exited the domain to the west in less than 15 days. Calculation of footprints according to equation (5) and coupling of transport to surface fluxes (equation (6)) were carried out offline. Footprints were gridded at a maximum resolution of 1/4°lon × 1/6°lat, corresponding to roughly 20 × 20 km^2. The horizontal size of the grid cells resolving the footprint was dynamically adjusted as the size of the footprint area increased, reducing computational time as well as preventing undersampling of surface fluxes at times when particles are distributed over extensive areas with large gaps between neighboring particles. The grid cell
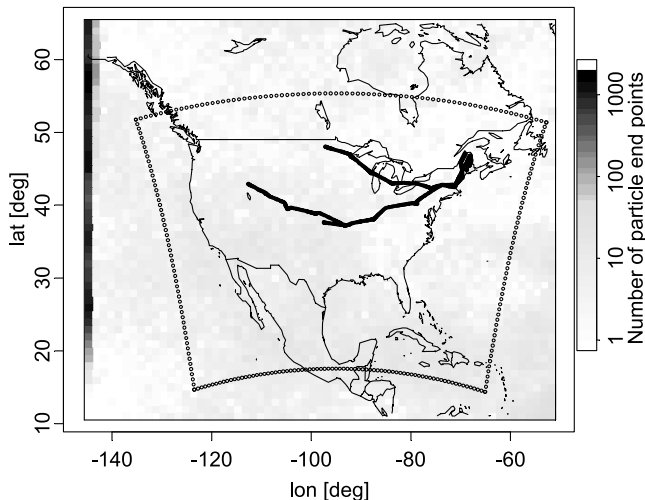
Figure 2. ROAM domain covering most of North America. Open circles indicate the subdomain for which EDAS winds where used. The starting locations for the particles along the flight track are indicated by the thick black lines. The gray scale indicates the number of particles ending in a particular grid cell after 15 days, or leaving the domain to the west.

resolution was degraded by keeping the number of grid elements representing the footprint between 18 and 36 as the footprint area increased. All surface flux grids (sections 2.3 and 2.4) were accordingly regridded before coupling to the footprints.

2.1.3. Convective Cloud Transport

[21] Vertical transport due to deep and shallow convection plays an important role in exchanging air between the mixed layer and the free troposphere. Unfortunately, archived meteorological fields currently lack subgrid convective fluxes from the original data assimilation model. The amount of convective transport on unresolved scales (below the EDAS 80 km resolution here) is likely to regulate the time τ_{exch} to replace the column of mixed-layer air by air from the mid troposphere. If convective transport is neglected, τ_{exch} is overestimated, causing an excessive build up of surface emissions in the lower atmosphere of the model.

[22] We developed a simple parameterization of convective cloud transport to give an upper estimate of the effect of subgrid-scale vertical redistribution and provide a lower limit for τ_{exch}. This estimate would then be compared with estimates with no convective cloud transport in order to bound the effects due to subgrid-scale vertical redistribution.

[23] The convective cloud transport scheme randomly assigns a vertical position between the surface and the limit of convection altitude (z_{LOC}) to each particle with a vertical position below z_{LOC}. The random redistribution is weighted by density and takes place whenever a new update of meteorological data is read (every 3 hours for EDAS) and whenever z_{LOC} is higher than z_i (mixed-layer height). z_{LOC} is taken to be the uppermost level at which the cloud air parcel is buoyant, i.e., the highest altitude where the parcel's virtual temperature exceeds that of the surrounding air. The virtual temperature profile in the cloud was calculated by

lifting a cloud parcel starting at the lowest model layer, lifted dry adiabatically to the lifting condensation level (LCL), and then moist adiabatically to the profile top. The presence of liquid water was neglected. We assumed that the virtual temperature from the assimilated field could be used as proxy for virtual temperature in the cloud-free environment at the levels at which cloud and non-cloud virtual temperatures are equal (i.e., z_{LOC}).

[24] The random redistribution scheme assumes that convective fluxes (both updrafts and downdrafts) are large enough to leave a perfectly well mixed column behind after each convective event, and that the convective events take place every 3 hours in grid cells with at least some convective available potential energy (CAPE). This simple approach represents an upper bound for the vertical redistribution caused by convective cloud transport, since any further vertical redistribution would not alter the particle distribution. We also conducted simulations without particle redistribution due to convection to provide a lower bound for convective transport effects. Throughout the rest of the paper, we will present model results from these two cases, with upper-limit and zero convective transport ("convective" and "non-convective" cases, respectively).

2.2. Determination of the Vegetation Signal

[25] Constraining regional/continental-scale exchanges of carbon between the biosphere and the atmosphere requires isolating the biospheric signal ($\Delta CO_{2,veg}$) from the other influences on the observed CO$_2$ ($CO_{2,meas}$), i.e., combustion signal ($\Delta CO_{2,comb}$) and advected background contributions ($CO_{2,bg}$; second term on the RHS of equation (1)):

$$\Delta CO_{2,veg} = CO_{2,meas} - \Delta CO_{2,comb} - CO_{2,bg} \qquad (7)$$

[26] The advected background mixing ratio is given directly by values calculated in ROAM, based on a lateral boundary condition (see section 2.2.1). The combustion signal $\Delta CO_{2,comb}$ consists of emissions from fossil fuel and biomass burning:

$$\Delta CO_{2,comb} = \Delta CO_{2,ff} + \Delta CO_{2,bb} \qquad (8)$$

[27] Because of the large forest fires in the domain in August 2000, we took an empirical approach to determining $\Delta CO_{2,comb}$, using the observed tracers instead of directly using CO$_2$ fossil fuel inventories. Further, the combustion signals resulted from emissions exhibiting a strong spatial variation, with a few localized populated areas causing most of the flux. This causes inaccuracies in transport or in spatial representation of emissions (on a coarse horizontal grid) to propagate into modeled combustion signals. By using direct measurements of a tracer indicative of combustion processes (CO), and using only relative emission ratios with less spatial variability (due to co-emission of the tracer with CO$_2$), we were able to make a more accurate determination of the combustion CO$_2$ signal actually encountered in flight. The excess of measured CO over background is an excellent tracer for biomass burning as well as for anthropogenic emissions [*Potosnak et al.*, 1999].

[28] We formulated the CO$_2$ combustion signal as a product of these enhancements in CO with CO$_2$:CO emis-

sion ratios from either fossil fuel emission inventories or from measurements in biomass burning plumes:

$$\Delta CO_{2,comb} = \Delta CO_{ff} \cdot \left(\frac{\Delta CO_{2,ff,inv}}{\Delta CO_{ff,inv}} \right) + \Delta CO_{bb} \cdot \left(\frac{\Delta CO_{2,bb}}{\Delta CO_{bb}} \right)$$

(9)

Here $\Delta CO_{2,ff,inv}$ and $\Delta CO_{ff,inv}$ are signals at the measurement location due to fossil fuel emissions inside the domain, obtained by coupling the transport model to emission inventories (see section 2.2.2).

[29] Excess CO is assumed to be due solely to combustion-derived emissions within the model domain (ΔCO_{comb}) and arises from the sum of fossil fuel emission (ΔCO_{ff}) and biomass burning (ΔCO_{bb}). However, this approach requires separation of biomass burning from fossil fuel sources; in section 4.2 we make use of the circumstance that during COBRA, biomass burning sources as well as their signatures were spatially separated from fossil fuel fluxes and associated plumes. ΔCO_{comb} can be derived from the observed CO and the background advected from the lateral boundary, corrected for estimated chemical production/loss:

$$\Delta CO_{comb} = \Delta CO_{ff} + \Delta CO_{bb} = CO_{meas} - (CO_{bg} - \Delta CO_{OH})$$

(10)

2.2.1. Lateral Tracer Boundary Condition

[30] The lateral tracer boundary condition is necessary to connect the regional tracer simulations to the global background tracer distribution and can be derived in two principal ways: (*a*) extraction of distributions of CO and CO₂ from a global model incorporating sources and sinks for these tracers (*b*) statistical characterization of the spatial and temporal dependence of tracer variations derived from available observations, keyed to contemporaneous measurements from surface stations. We decided on the statistical approach because of large discrepancies between global models in simulating the CO₂ distribution in the upper troposphere [*Law et al.*, 1996] and the fact that our study concentrates on sources/sinks in a limited domain instead of quantifying global-scale fluxes. The western boundary condition was selected to be at 145 W that, because of the dominant westerly flow over the United States, represents the tracer concentrations of air parcels over the ocean before being affected by terrestrial sources and sinks. Indeed, most of the particles (about 66%) cross the 145 W after being transported for ~6 days backward from the receptor (Figure 2), and most of the remainder (20%) reside inside the ROAM domain. Only small numbers exit the domain to the north (6%), east (4%), or south (4%). The initial value (concentration of CO₂ and CO) for each particle was then calculated on the basis of the latitude, altitude, and time when it crosses the western boundary or when the calculation was stopped after 15 days. This approach neglects zonal gradients in the global tracer distribution for particles not leaving the domain to the west. However, zonal gradients are much smaller than meridional gradients; e.g., for CO₂ the difference between samples taken at Arctic stations Alert (62 W) and at Barrow (156 W) are less than 1 ppm during the month of August 2000; similarly the differences between subtropical stations Kumakahi (155 W) and Ber-

muda (65 W) were less than 0.5 ppm. In contrast, meridional differences between the Arctic and subtropics were much larger, at ~6 ppm.

[31] Time-dependent tracer fields for CO₂ and CO as functions of altitude and latitude over the Pacific were generated by statistical analyses of ground-based and airborne measurements of CO and CO₂ for the past 20 years (see Table 2 in the companion paper [*Gerbig et al.*, 2003]). The analysis was done in two steps: first, mixing ratios at the ground were represented by an analytical function of time derived from measurements at selected ground stations, and then this function was propagated vertically to fit the airborne measurements at higher altitudes.

[32] To retrieve continuous tracer time-series at the surface, the measurements at three selected ground stations (Cape Kumakahi (HI), Cold Bay (AK), and Barrow (AK)) for the 20 year period starting 1/1/1980 (in case of CO₂) and for the 10 year period starting 5/26/1992 (in case of CO) were represented by its Fourier components, with a cut-off of 5 days. The GLOBALVIEW data set [*GLOBALVIEW-CO₂*, 2002] (also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW) was used to fill gaps in the CO₂ time series before calculating the Fourier components. CO data were provided by P. Novelli (personal communication, 2001) [*Novelli et al.*, 1998], and missing values were filled by linear interpolation. The result of the Fourier analysis was an analytical representation of the time-series for CO₂ and CO, closely resembling the measurements. These functions were linearly interpolated in latitude to provide surface tracer mixing ratios $CO_{2,S}(t, lat)$ and $CO_S(t, lat)$ at all required times t and latitudes lat.

[33] To represent upper air measurements, airborne and station data collected above 300 m over the northern Pacific (see Table 2 in the companion paper [*Gerbig et al.*, 2003]) were merged. We also included CO data from Niwot Ridge, Colorado and continental flask data collected at 8 km above Carr, Colorado, for CO₂ and CO. Aircraft data with potential influence from the stratosphere were excluded using a filter based on O₃, N₂O, CO, potential temperature, H₂O, and altitude. Flask data were given a 10 fold higher weight to account for the different nature of the data compared to in-situ measurements; the 1-minute averages of the in-situ data are assumed to be correlated within a timescale of 10 min (assumed average duration of an atmospheric layer when sampled by an aircraft flying mostly horizontally).

[34] We used a Green's function to fit the upper air measurements for CO₂ by vertically propagating the surface tracer mixing ratios [cf. *Andrews et al.*, 1999]. The Green's function makes use of the fact that CO₂ is conserved in the troposphere. In the companion paper we calculated the monthly mean vertical profiles, based on a comparison between the marine boundary layer reference from the GLOBALVIEW data set and the extensive CO₂ measurements (Figure 4 in the companion paper [*Gerbig et al.*, 2003]). A simple Green's function approach is expected to capture much of the seasonality of vertical gradients. A simpler statistical model was fitted for CO, which is not conserved, involving an average CO gradient, an altitude-dependent damping of high-frequency variability, and an altitude-dependent lag time. Details of the fitting procedures are given in Appendix A.
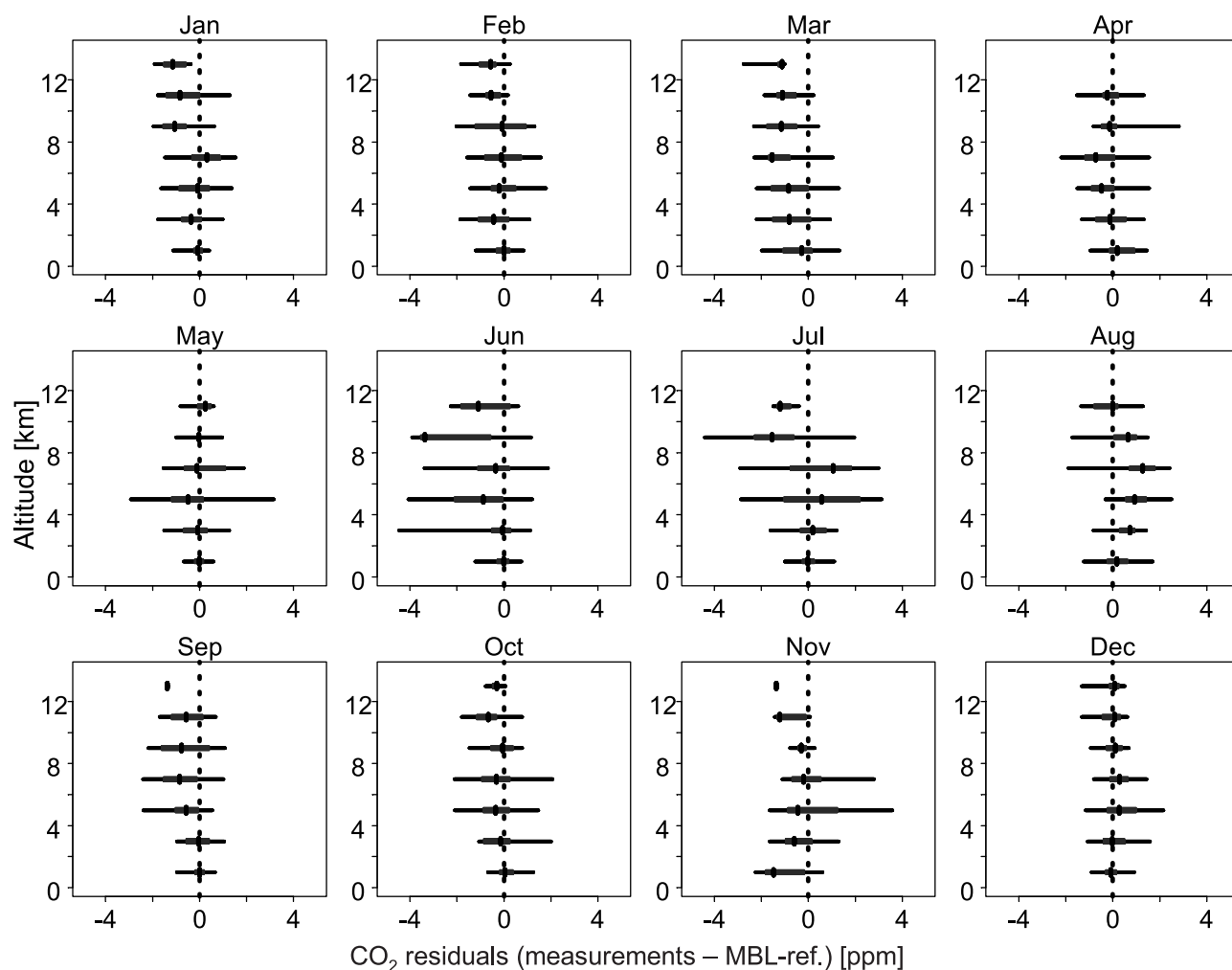
Figure 3. Distribution of residuals between measured CO$_2$ over the Pacific and Green's-function-derived CO$_2$ (measurements - Green's function) in ppm (x-axes) for different altitude intervals (y-axes) and for each month. Boxes indicate the central 50%, horizontal lines the central 90%, and vertical bars the median.

[35] Meridional cross sections for CO$_2$ and CO were generated at a resolution of 0.5 km altitude by 2.5° latitude and with a time resolution of one day, using the parameterizations derived in Appendix A (equations (A1) and (A4)). To assess the error of these fields, cross sections were calculated on a weekly basis for the 20 year period in the case of CO$_2$, and for the 10 year period in case of CO, and merged with the observations. Residuals in CO$_2$ (measurements - statistical model) have a mean of 0.22 ppm, and a standard deviation of 1.15 ppm (1σ). The seasonal bias error at elevated altitudes as found for the MBL-reference CO$_2$ (Figure 3; compare Figure 6 in the companion paper) is greatly reduced. Small biases of up to 1 ppm are observed in August and September, with opposite signs. This is likely due to the fact that seasonality in vertical transport is only represented by a sinusoidal amplitude modulation in the timescales (see Appendix A), whereas seasonal changes in transport patterns are likely to occur more abruptly.

[36] Residuals for CO have a mean of 1.2 ppb, and a standard deviation of 22 ppb. Most of the upper atmosphere data used here were collected in campaigns, so it is possible that there is sampling bias due to the specific objectives of these campaigns. It is difficult, however, to assess how such a bias would affect the analysis: Large variability for CO could be caused by selective sampling of polluted layers; it is also possible that mid-tropospheric air is more influenced by proximate pollution than the marine boundary layer, thus causing higher and more variable CO. Indeed, using the mean CO$_2$:CO ratio from fossil fuel emissions (0.03 ppm/ppb, see section 4.2), the 22 ppb variability for CO residuals corresponds to only 0.66 ppm of CO$_2$, smaller than the observed variability in residuals for CO$_2$.

[37] The particle duration of 15 days is not negligible compared to an average photochemical lifetime of 30 days for CO in summer. Hence both photochemical loss due to reaction with OH and production from CH$_4$ oxidation were implemented in a simplified way. Climatological values for OH at the mean position (latitude and altitude) of a particle ensemble were adopted from the GEOS-CHEM OH fields (L. Jaegle, http://marzipan.atmos.washington.edu/ion_script/GEOS/main_geosoh.html), and rate constants for reactions of CO with OH and CH$_4$ with OH [*Atkinson et al.*, 1997] were calculated using pressure and temperature measured at the aircraft location. Oxidation of

Table 1. AmeriFlux Sites Used for Parameterization of NEE for Different Vegetation Types[a]

Site	Vegetation Type (IGBP Classification)	Vegetation (AmeriFlux Description)	Vegetation Type (Simplified)	$\sigma(\lambda_R)$	$\sigma(\lambda_{GEE})$
Duke Forest (NC)	evergreen needleleaf forest	even-aged loblolly pine plantation	forest	0.66	1.1
Harvard Forest (MA)	deciduous broadleaf forest	temperate deciduous forest	forest	0.66	1.1
Blodgett Forest (CA)	evergreen needleleaf forest	mixed evergreen coniferous forest dominated by ponderosa pine	forest	0.66	1.1
BOREAS NSA - Old Black Spruce (Canada)	evergreen needleleaf forest	evergreen coniferous forests, boreal needle forest, old black spruce trees	forest	0.66	1.1
Gainesville (Austin Cary) (FL)	evergreen needleleaf forest	65 yr regenerating slash/longleaf pine	forest	0.66	1.1
Metolius Research Natural Area - young ponderosa pine (OR)	evergreen needleleaf forest	intermediate (56 yrs) ponderosa pine	forest	0.66	1.1
Niwot Ridge Forest (Co)	evergreen needleleaf forest	subalpine coniferous forest	forest	0.66	1.1
Univ. of Mich. Biological Station, (MI)	mixed forest	mid-aged conifer and deciduous	forest	0.66	1.1
Willow Creek (WI)	deciduous broadleaf forest	sugar maple-basswood forest, with some green ash and red oak	forest	0.66	1.1
WLEF Wisconsin Tall Tower (WI)	mixed forest	temperate/boreal forest, lowland and wetland forest	forest	0.66	1.1
Lethbridge (Canada)	grasslands	short/mixed grass prairie (C3/C4)	shrubland	20.2	3.9
Barrow (AK)	barren or sparsely vegetated	arctic tundra	shrubland	20.2	3.9
Bondville (IL)	croplands	annual rotation between corn (C4) - 2001, soybeans (C3) - 2000	cropland	0.27	0.15

[a]The last two columns show the uncertainty of the scaling factors as derived from 3-day integrated residuals (see text).

CH_4 was assumed to yield one CO per oxidized CH_4. The calculated net loss for CO was small (about 14 ppb on average, with a maximum of 30 ppb), but omitting it would give biased results. No chemical production or loss was implemented for emission-derived CO (next paragraph), since most of the signal resulted from emissions close to the receptor location.

2.2.2. Fossil Fuel Inventories

[38] Emissions of CO_2 from fossil fuel, cement production and gas flaring were taken from the $1° \times 1°$ inventory compiled by *Marland et al.* [1997], which is the 1995 version of the *Andres et al.* [1996] inventory. To account for changes in emissions between 1995 and 2000, a 10% increase was assumed by propagating the trend between 1992 and 1996 to 2000; because of the lower trends of the years 1997–1999 of less than 1% [*Environmental Protection Agency* (*EPA*), 2001] this increase is likely to be high by a few percent.

[39] Surface fluxes of CO were calculated from a combination of the NAPAP 1990 inventory for the northeastern United States ($1/6°$Lat. $\times 1/4°$Lon.) [*EPA*, 1993] and the GEIA inventory ($1°$Lat. $\times 1°$Lon.) [*Benkovitz et al.*, 1996]. Since the GEIA inventory does not include emissions for CO, they were calculated from the low-level NO_x emissions using a linear regression between NAPAP CO emissions and the GEIA low-level NO_x emissions in the overlapping area (CO = $12.5 \times NOx$(GEIA,low); $R^2 = 0.84$). Time-of-day and day-of-week factors were applied to account for time-dependence of emission fluxes [*Ebel et al.*, 1997]. Since no information on the diurnal and weekly variations are given by the CO_2 fossil fuel emission inventories, the time factors for CO emissions were also used for the CO_2 emissions; however, the amplitude of these variations was reduced by a factor of 2.5 in rough agreement with diurnal variations of anthropogenic CO_2 emissions over Japan [*Kondo et al.*, 2001].

2.3. Biosphere Flux Model

[40] The main role of the biosphere flux model in the analysis framework is to provide a surface boundary condition for CO_2, which captures the dominant patterns of spatial and temporal variability, i.e., vegetation coverage and response to environmental drivers (light and temperature). Terrestrial fluxes of CO_2 were modeled by linearly scaling eddy covariance observations of net ecosystem exchange (NEE) from the AmeriFlux network [*Baldocchi et al.*, 2001] for different vegetation types (denoted by i), projected regionally using land-cover data:

$$NEE_i = \lambda_{i,R} \cdot R_i + \lambda_{i,GEE} \cdot GEE_i \quad (11a)$$

$$R_i = \beta_i T; \quad GEE_i = \frac{a_i \cdot SWRF}{b_i + SWRF} \quad (11b)$$

[41] Here $\lambda_{i,R}$ and $\lambda_{i,GEE}$ are scaling factors (see below), R_i is the respiration flux, GEE_i is the gross ecosystem exchange, T is the atmospheric temperature in °C at 2 m; $SWRF$ is the downward short wave radiative flux from the assimilated meteorological fields (used as a proxy for photosynthetically active radiation). The parameters β_i (temperature coefficient for respiration flux, autotrophic + heterotrophic), a_i (maximum GEE), and b_i (with a_i/b_i as the quantum efficiency) were obtained from least-square fits to the hourly AmeriFlux data (available under http://public.ornl.gov/ameriflux) for each vegetation class over the months July and August in 2000 (Table 1). The model explained most of the hourly variance for the crop site ($r^2 = 0.79$) and for forests ($r^2 = 0.56$).

[42] The IGBP 1-km resolution vegetation data [*Belward et al.*, 1999] were regridded to the ROAM domain at the different horizontal resolutions starting at $1/6°$Lat. $\times 1/4°$Lon., to give relative coverage for each vegetation type over the ROAM domain. Since tower flux data are not available for each of the 17 different vegetation classes, the vegetation classification was simplified into 5 classes: forests, shrublands (open and closed shrublands, savannas, grassland, and barren or sparsely vegetated), croplands (croplands and cropland/natural vegetation mosaic), wet-

lands, and water bodies. The scaling factors $\lambda_{i,R}$ and $\lambda_{i,GEE}$ then account for mismatches in the classification and upscaling from flux-site to the larger regions. These scaling factors, the adjustable parameters for the biospheric CO_2 flux, will be estimated using a Bayesian inversion described in section 4.3.

[43] The net CO_2 flux was assumed to be negligible for water bodies and wetlands. The upper limit of the air-sea fluxes was estimated to be $\sim -0.1\mu moles/m^2/s$ (uptake) on the basis of recent pCO_2 data [*Lefevre et al.*, 1999]. Hence air-sea exchange fluxes are smaller than terrestrial fluxes by 3 orders of magnitude and can be neglected even though a significant part of the ROAM domain is covered by ocean. Wetlands in most of the domain tend to be very small units that cannot be distinguished from the surrounding forests.

[44] Prior estimates for the different scaling factors were all set to 1.0, and prior uncertainties of these scaling factors were estimated from the residuals between hourly eddy covariance data and hourly predictions using equation (11) (with scaling factors set to one). Ideally, we would derive the covariance matrix from residuals in R and GEE, taking into account the spatial and temporal correlations between different times and locations "seen" by the airborne measurements (the footprint elements in equation (5)). Unfortunately there is not enough spatial coverage in the AmeriFlux data to allow for estimation of the spatial correlation in the residuals, but we can make use of the temporal coverage at the different sites. We assume that the biospheric signals "seen" at the receptors relate to 3 day aggregates of R and GEE in the footprint region, corresponding to an average turnover time between PBL and free troposphere. To derive the uncertainty in the scaling factor for respiration, we integrated the nighttime residuals (hourly flux measurements − hourly predictions) over 3 days and divided by the 3-day integrated nighttime NEE (predicted) for each site. For each flux site this results in about 20 ratios for the two-month period. The prior uncertainty in $\lambda_{i,R}$ was then calculated as the standard deviation of these ratios for a given vegetation class i (see Table 1). Similarly, for GEE we took the ratio of 3-day integrated daytime residuals in NEE to 3-day integrated daytime NEE (predicted). If the spatial coverage would allow for derivation of the spatial correlation in these residuals, the prior uncertainty in the scaling factor over a large area (such as the footprint of the measurements, see section 4.1) would likely be less than our estimate since residuals are likely to be correlated over a length scale smaller than that covered by the area and leading to cancellation of errors. In this sense we are using a conservative estimate for the prior uncertainty.

[45] For forests the uncertainties in $\lambda_{i,R}$ and $\lambda_{i,GEE}$ are on the order of 1 (Table 1), corresponding to a signal to noise of 1 for 3-day aggregated fluxes. For shrubland the values are significantly larger, resulting from the diversity in the two AmeriFlux sites (short/mixed grass prairie and arctic tundra) assumed to be representative for this vegetation class. For cropland values are smaller, reflecting the fact that only a single site was used. In this case residuals do not include a spatial prediction error, so we arbitrarily increased the uncertainty by a factor of 10 (reducing the weight of the prior value) to account for the fact that a single flux site over a crop field is not representative for the diverse crop types and management regimes found within this vegetation class.

2.4. Optimization Approach

[46] In order to link the large-scale constraints from the vegetation signal $\Delta CO_{2,veg}$ to local-scale information provided by the eddy flux measurements, we scaled up the tower fluxes to the continent using the vegetation classification described in the previous section and equations (6) and (11) to calculate the resulting vegetation signals at the aircraft location. The scaling factors used to scale up and adjust biospheric fluxes constrained by eddy flux measurements were calculated from a formal Bayesian inversion approach that incorporates a priori estimates and their associated errors [*Rodgers*, 2000].

[47] The scaling factors for gross exchange from different vegetation classes in equation (11) are related to the measurements (the vegetation signals $\Delta CO_{2,veg}$) by the following equation (we mostly follow the notation of *Rodgers* [2000]):

$$\mathbf{y} = \mathbf{K}\boldsymbol{\lambda} + \boldsymbol{\varepsilon} \quad (12)$$

[48] Here $\mathbf{y}$ represents a vector of measurements, $\mathbf{K}$ is the Jacobian matrix relating the measurement vector to the state vector, $\boldsymbol{\lambda}$ is the state vector of scaling factors, and $\boldsymbol{\varepsilon}$ is an error vector accounting for uncertainty in the measurements and in the modeling framework.

[49] The measurement vector $\mathbf{y}$ has one element for each of the different receptors j at location $\mathbf{x}_{rj}$, and at time t_{rj}: $\mathbf{y} = [\Delta CO_{2,veg}(\mathbf{x}_{r1},t_{r1}), \ \Delta CO_{2,veg}(\mathbf{x}_{r1},t_{r1}), \ \ldots]^T$ (where ^{T} indicates the transpose). The Jacobian matrix $\mathbf{K}$ has as elements the advected signals after equation (6), where fluxes F are taken from the prior gross fluxes (photosynthesis GEE and respiration R) for the different vegetation classes using equation (11b) along with temperature and shortwave radiation from EDAS:

$$\mathbf{K} = \begin{bmatrix} \sum_i f(\mathbf{x}_r,t_r|\mathbf{x}_i,t_i) \cdot GEE_f(\mathbf{x}_i,t_i) & \sum_i f(\mathbf{x}_r,t_r|\mathbf{x}_i,t_i) \cdot R_f(\mathbf{x}_i,t_i) & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \quad (13)$$

The state vector $\boldsymbol{\lambda}$ combines the scaling factors for the gross fluxes from different vegetation classes, $\boldsymbol{\lambda} = [\lambda_{GEE,forest}, \lambda_{R,forest}, \lambda_{GEE,crop}, \lambda_{R,crop}, \ldots]^T$. Optimum posterior estimates of scaling factors are obtained by minimizing the cost function J, for which we chose the standard least squares formulation:

$$J(\boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{K}\boldsymbol{\lambda})^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{K}\boldsymbol{\lambda}) + (\boldsymbol{\lambda} - \boldsymbol{\lambda}_{prior})^T \mathbf{S}_{prior}^{-1} (\boldsymbol{\lambda} - \boldsymbol{\lambda}_{prior}) \quad (14)$$

[50] Here $\mathbf{S}_\varepsilon$ and $\mathbf{S}_{prior}$ are the error covariance matrix for the vegetation signals and for the prior scaling factors, respectively. $\mathbf{S}_{prior}$ is derived from comparison between fluxes from the biospheric model and observed eddy covariance fluxes (section 2.3, Table 1). Minimizing J results in posterior estimates for $\boldsymbol{\lambda}$ consistent with both the measurements along the flight track and the prior estimates for gross fluxes, given the respective uncertainties $\mathbf{S}_\varepsilon$ and $\mathbf{S}_{prior}$:

$$\widehat{\boldsymbol{\lambda}} = \left(\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{prior}^{-1} \right)^{-1} \left(\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{y} + \mathbf{S}_{prior}^{-1} \boldsymbol{\lambda}_{prior} \right) \quad (15)$$

[51] The posterior uncertainty of $\hat{\lambda}$ is expressed in form of the error covariance matrix $\hat{\mathbf{S}}_{\lambda}$:

$$\hat{\mathbf{S}}_{\lambda} = \left(\mathbf{K}^T \mathbf{S}_{\varepsilon}^{-1} \mathbf{K} + \mathbf{S}_{prior}^{-1}\right)^{-1} \quad (16)$$

[52] Proper treatment of the uncertainties in the measurements and various parts of the model is important, as these uncertainties represent weighting factors for the optimization. Throughout the following sections we provide estimates for these uncertainties at the different measurement locations. The error covariance matrix $\mathbf{S}_{\varepsilon}$ is a sum of the error covariance in the vegetation signal ($\mathbf{S}_{veg}$, see section 4.2) and different components describing errors in the modeling framework:

$$\mathbf{S}_{\varepsilon} = \mathbf{S}_{veg} + \mathbf{S}_{part} + \mathbf{S}_{eddy} + \mathbf{S}_{transp} + \mathbf{S}_{aggr} + \mathbf{S}_{ocean} \quad (17)$$

[53] Here $\mathbf{S}_{part}$ is the random error due to particle statistics (section 3.1), $\mathbf{S}_{eddy}$ is the error due to unresolved eddies, $\mathbf{S}_{transp}$ describes the error in the mixed-layer height in the transport model, $\mathbf{S}_{aggr}$ represents the error due to aggregation of fluxes into large spatial regions (the footprints of the northern and southern surveys), and $\mathbf{S}_{ocean}$ accounts for neglecting oceanic fluxes in the inversion.

[54] For $\mathbf{S}_{part}$ we used a diagonal matrix (uncorrelated errors). The square roots of the diagonal elements were assumed to be 13% of the sum of a typical mixed-layer signal, in accordance with the relative uncertainty for mixed-layer receptors derived in section 3.1. For $\mathbf{S}_{eddy}$ we took values for the standard deviation of CO$_2$ within each missed layer profile as described in the companion paper, ranging from 0.2 to 2 ppm. A transport error $\mathbf{S}_{transp}$ was included to account for the mismatch in modeled and measurement derived mixed-layer height z_i. The coarse resolution in the analyzed fields (250 m close to the surface) and the difficulties in deriving z_i from analyzed profiles contribute to errors in z_i with residuals between measured and modeled z_i averaging 600 m (standard deviation).

[55] The square roots of the diagonal elements of $\mathbf{S}_{transp}$ were calculating by multiplying the relative error in z_i (represented by $(z_i(\text{meas})-z_i(\text{mod})/z_i(\text{mod}))$ with 10 ppm as an average vegetation signal in the PBL; here $z_i(\text{meas})$ corresponds to z_i estimates from the tracer profiles [see *Lin et al.*, 2003]. This error affects only measurements at altitudes below the measured or modeled mixed-layer height; average values are around 3 ppm. The uncertainty due to improper parameterization of convective fluxes is not included as an error term in equation (17); instead we explicitly use the separate transport simulations corresponding to minimal convection (without subgrid convection) and maximum convection (with subgrid convection as described in section 2.1.3. For the aggregation error $\mathbf{S}_{aggr}$, we chose diagonal elements of $\mathbf{S}_{aggr}$ of $(3\ \text{ppm})^2$. As we will show in section 4.4, this error corresponds to aggregation scales of ~ 2000 km (the approximate size of the footprints).

[56] In order to obtain a vegetation signal representing continental CO$_2$ exchange, the signal from oceanic fluxes should be subtracted from the vegetation signal in equation (7). However, using the upper limit flux for uptake of 0.1 μmol/m^2/s [*Lefevre et al.*, 1999] this "correction" corresponds to less than 2 % of the vegetation signal. Therefore we have omitted this correction, but instead included an additional error term $\mathbf{S}_{ocean}$ for the vegetation signals corresponding to the upper limit flux.

[57] The transport error, the aggregation error as well as the error from neglecting oceanic fluxes are assumed to be correlated within individual profiles. This was ensured by using exponentially decaying covariances with increasing temporal and spatial distance between receptors, with a decorrelation length of 10 km and timescale of 12 min (corresponding to the spatial and temporal scale of the aircraft path during a profile through the mixed layer).

3. Transport Model Tests

3.1. Dependence on Particle Number and Reproducibility

[58] One would ideally use a sufficiently large number of particles (N_{tot}) to represent the ensemble properties of the transport to a given measurement location. Since model run time increases proportionally to the number of particles used, we examined the dependence of the "sampling" error on N_{tot} in an attempt to minimize the run time. We examined the convergence as a function of N_{tot} of the simulated mixed-layer signal by coupling the footprints with biosphere-atmosphere exchange fluxes (i.e., by combining equations (6) and (11)) for a subset of the receptors (we chose five different mixed-layer measurement locations along the northern survey). Particle number was varied between 50 and 1000 for each measurement location, and the model was run 100 times for each location and N_{tot} to derive the statistical variance (i.e., the "sampling" error). The standard deviation obeyed Poisson statistics, with stdev(CO$_2$)/CO$_2 \sim$ sqrt(N_{tot}), as expected. For 100 particles the error due to the stochastic nature of the model was 13%. The resulting computational cost was acceptable (3 min per receptor location for a 15-day simulation without convection for a 600 MHz processor running Linux, and twice the time with convection). The stochastic nature of this error and the complete lack of correlation between different receptors allow for proper representation of this error in the ROAM framework as a random, uncorrelated error.

3.2. Depth of "Surface Layer"

[59] The dependence of model results (vegetation CO$_2$ signals) on the initial column height h (see equation (2)) used to dilute surface fluxes at each time step was investigated by varying h between $0.1 \cdot z_i$ and $1.0 \cdot z_i$. No significant change in the modeled vegetation signal was found. However, at shallow initial column depths the number of particles influenced by surface fluxes during a time step is smaller, which causes the statistical noise to increase. We chose a depth of $0.5 \cdot z_i$ as the column height for which timescales for vertical mixing roughly matches the model timestep for advection.

4. Results

[60] The analysis framework described in section 3 will now be applied to the COBRA data to derive constraints for regional fluxes (Figure 1). We first present the footprints,
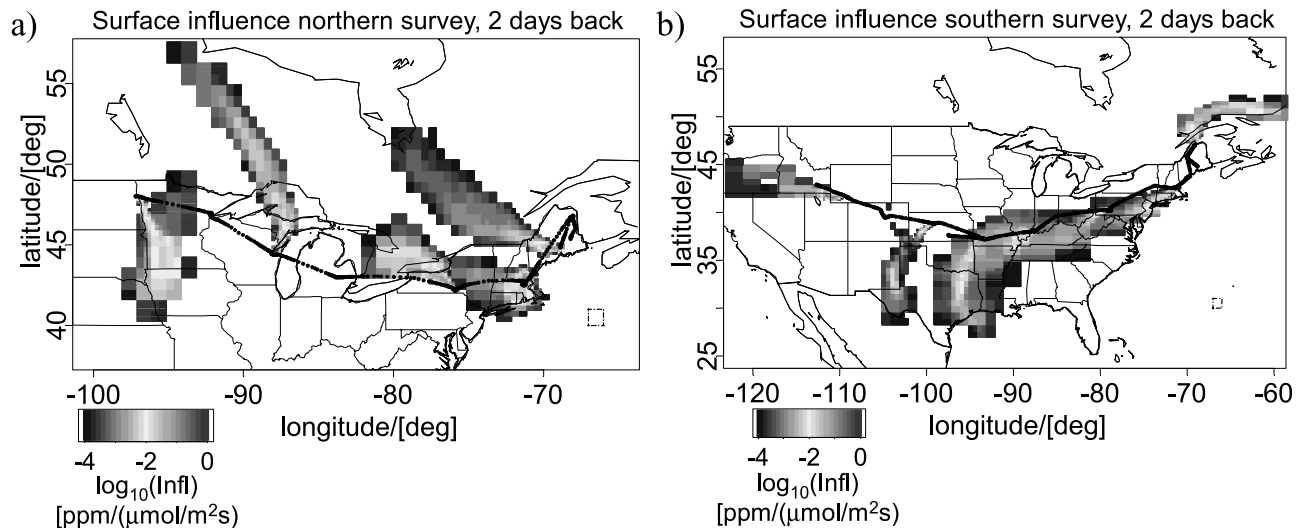
Figure 4. Footprints for mixed-layer receptors, integrated over the last 2 days prior to the measurements, calculated for the non-convective case for the (a) northern and (b) southern surveys. The different pixel sizes are a result of the dynamic grid resolution. The square with an edge length of 1 degree over the Atlantic Ocean is shown as reference. See color version of this figure at back of this issue.

i.e., the sensitivities of mixing ratios measured along the flight track to upstream surface fluxes (section 4.1). Then we combine measured CO_2 with advected background mixing ratios for CO_2 and an estimate of the fossil fuel contribution to CO_2 to isolate the CO_2 vegetation signal from the observations (section 4.2). In section 4.3 we use CO_2 signals due to biospheric fluxes according to equation (11) as constraints to optimize scaling factors that are consistent with the observation-based CO_2 vegetation signal, and estimate regional flux distributions and their response to meteorological conditions.

4.1. Footprints for Mixed-Layer Receptors

[61] To visualize the surface influence and its spatial variation, we calculated time integrals of footprints over the 2 days prior to the measurement time, when surface influence is strongest. For this illustration we chose a small subset of receptor locations along the flight track of both surveys, at altitudes below mixed-layer top (z_i). The time integrated footprint (Figure 4) shows the signal in ppm found at the receptor, caused by a flux of 1 μmol m^{-2} s^{-1} from each surface location. Footprints vary on small spatial scales, especially proximate to the receptor. Farther from the receptors, the dynamic grid resolution increased the grid size to about 1 degree. The footprints reveal a strong contrast between the northern and the southern survey, with north-westerly influence for the north, and south-westerly influence for the south. By comparing the time-integrated footprints in Figure 4 with the vegetation condition index [*Kogan*, 1997] (Figure 2e in the companion paper), it becomes clear that the mixed layer in the northern survey was influenced by vegetation with fair to favorable growing conditions, while the southern survey mixed-layer air, especially in the western part, was influenced by stressed vegetation. This is consistent with the large differences in the observed CO_2, with a significant draw down in the lower atmosphere along the northern transect and CO_2 emission in

the southern transect (Figures 2a and 2c in the companion paper).

[62] The time evolution of surface influence is shown by integrating footprints over the surface of the entire model domain, averaged over time periods ranging from 12 hours for the first 3 days back, to 72 hours for 9 to 5 days back (Figure 5). The values displayed represent signals resulting from a flux of 1 μmol m^{-2} s^{-1} acting over 24 hours. They start at values around 1.5 ppm per μmol m^{-2} s^{-1} and decay by factors of 2–3 over 48 hours. Since these are spatially integrated footprints, they are not affected by horizontal dispersion, but only by vertical exchange processes like subsidence and convective cloud transport, which exchange air between the mixed layer (ML) and the free troposphere (FT). For transit-times which are small compared to the timescale for exchange between ML and FT (τ_{exch}) one can simply calculate spatially integrated footprints by dividing the number of molecules emitted from the surface flux (here 1 μmol m^{-2} s^{-1}) during a day by the number of molecules in the column in contact with the surface, with a height corresponding to the maximum daily z_i (mixed-layer height). This gives an influence of 1.4 ppm/μmol m^{-2} s^{-1} for a z_i of 1.5 km, in agreement with the values for 0–12 hours prior to arrival in Figure 5. For transit times much longer than τ_{exch} one can assume a vertically well-mixed troposphere, for which spatially integrated footprints asymptotically approach a lower limit of 0.3 ppm/μmol m^{-2} s^{-1}. However, because of the limited model domain, particles start leaving the domain, causing the spatially integrated footprints derived from particle distributions to decrease below this limit.

[63] Figure 5 clearly shows that convection has a major impact on the decay of spatially integrated footprints with time: with convection, the influence decays to less than 1/e of the initial influence after a single day, while without convection the influence decays to 1/e only after ~5 days. These e-folding times correspond to the timescale for mixing between mixed layer and free troposphere. The
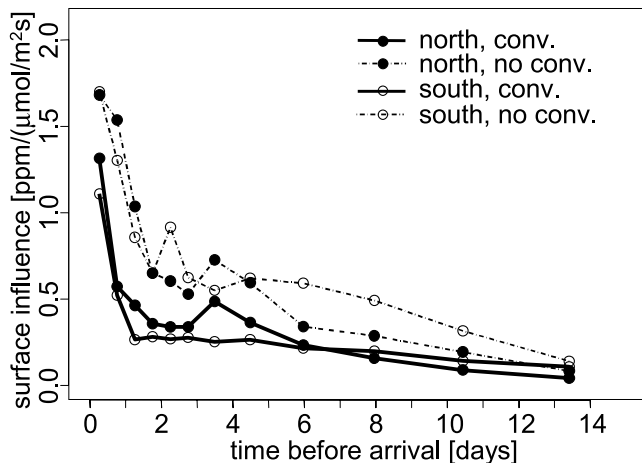
Figure 5. Spatially integrated surface influence for mixed-layer receptors as a function of time before arrival at the flight track, for the northern (solid symbols) and the southern transect (open symbols). Results for the convective case are shown as solid lines, and for the non-convective case as dashed lines.

3-day turnover time used in section 2.3 to derive the prior uncertainty from residuals thus represents a reasonable value for the COBRA period.

4.2. Observation-Based CO_2 Combustion and Vegetation Signals

[64] In this section we first derive CO_2 combustion signals from the measurement based CO combustion signal (ΔCO_{comb}) and the CO_2/CO emission ratio following equations (8)–(10), and then we calculate the measurement based vegetation signal ($\Delta CO_{2,veg}$, equation (7)) as well as the associated uncertainty (S_{veg}).

[65] ΔCO_{comb} (Figures 6a and 6b) shows values close to zero for most of the troposphere, except for polluted layers due to anthropogenic emissions and biomass burning, where CO concentrations are elevated by more than 100 ppb. Differences between the runs without convection (Figure 6a) and with convection (Figure 6b) are small, indicating that the advected background CO is similar for both cases. When comparing these "measured" combustion signals to CO signals from the transport model and the inventory ($\Delta CO_{ff,inv}$, Figures 6c and 6d), it is obvious that they differ significantly from each other, not only because of the lack of biomass burning emissions in the model but also because of inappropriate convective mixing. Without subgrid convection (Figure 6c) the model overestimates fossil fuel CO near Boston by a factor of 2, while the simulation incorporating subgrid convection (Figure 6d) exhibits closer agreement to measured values.

[66] The partitioning of the contribution of biomass burning versus anthropogenic sources to ΔCO_{comb} is required to derive $\Delta CO_{2,comb}$, as the two combustion sources yield different CO_2:CO emission ratios. Here it would be desirable to have another independent tracer to distinguish between the different sources, such as CH_4. However, methane was only measured from flask samples, with low spatial coverage. Fortunately, biomass burning sources were not co-located with anthropogenic sources, and the resulting

biomass burning signals in ΔCO_{comb} were not observed at the same locations as the anthropogenic pollution (Figures 6a and 6b). We thus could identify biomass burning events, following equation (10), as periods when the modeled $\Delta CO_{ff,inv}$ underestimates the "measured" CO combustion signal ΔCO_{comb} by more than a factor of 4, and the residual was assumed to represent the biomass burning signal $\Delta CO_{bb} = \Delta CO_{comb} - \Delta CO_{ff,inv}$. This underestimation-factor is somewhat arbitrary, and is only regarded as a preliminary method to select the few locations affected by biomass burning. This method is not expected to succeed in cases with co-located biomass burning and fossil fuel signals.

[67] In the absence of biomass burning ($\Delta CO_{bb} = 0$, observed for 88% of the receptors), the CO_2 combustion signal $\Delta CO_{2,comb}$ was derived after equation (9), using the CO combustion signal and the inventory-based emission ratio averaged over the footprint. For very low $\Delta CO_{ff,inv}$ (<1 ppb) an average emission ratio of 0.03 ppm/ppb was assumed to avoid instability in the emission ratio $\Delta CO_{2,ff,inv}:\Delta CO_{ff,inv}$, and negative values for ΔCO_{comb} (found for 2% of all measurements) were set to zero to avoid "negative" fossil fuel contributions. In the presence of CO from biomass burning (at about 12% of the receptors), fossil fuel emissions for CO_2 were adopted from the inventory-based $\Delta CO_{2,ff,inv}$, and biomass burning CO_2 was calculated as $\Delta CO_{2,bb} = (CO_2/CO)_{bb} \cdot \Delta CO_{bb}$, with the biomass burning emission ratio $(CO_2/CO)_{bb}$ estimated to be 0.0071 ppm/ppb from our observed correlations in biomass burning layers.

[68] Cross sections of $\Delta CO_{2,comb}$ (combustion signal including biomass burning) for the northern survey are presented in Figures 6e and 6f, for the cases with and without convection. They look similar, with relative differences much smaller than for inventory- based fossil fuel CO signals ($\Delta CO_{ff,inv}$, Figures 6c and 6d) or for inventory-based CO_2 fossil fuel signals ($\Delta CO_{2,ff,inv}$, not shown). Thus by using the measured CO enhancements and inventory-based emission ratio, we reduce biases in combustion-related CO_2 due to inaccuracies in transport (e.g., subgrid convection). The $\Delta CO_{2,comb}$ is influenced predominantly by emissions from the major cities of Boston and Denver with values of up to 6 ppm, while biomass burning contributed up to 2 ppm CO_2 in the free troposphere.

[69] Uncertainties for $\Delta CO_{2,comb}$ are estimated for each receptor location by propagating errors in the separate terms of equations (9) and (10). Uncertainties in ΔCO_{comb} are on average 23 ppb, arising from uncertainties in the boundary condition (22 ppb), in the assumed OH field (5 ppb for a 50% uncertainty in OH), and in the measurements (2%).

[70] Uncertainties in $\Delta CO_{2,comb}$ also require an estimate of the uncertainty in CO/CO_2 emission ratios. Only qualitative estimates are available for uncertainties of emission inventories on the grid scale; the relative errors are assumed to be significantly larger than country-level total emissions [*Marland et al.*, 1997]. However, because of common factors contributing to the uncertainties in both CO and CO_2 emissions (e.g., population statistics, fuel use, etc.), we believe that their errors are partly correlated. The consequence is that relative uncertainties in the emission ratio are less than the individual relative uncertainties. Lacking detailed information, we estimated the uncertainty in the emission ratio to 30%. In presence of biomass burning, we
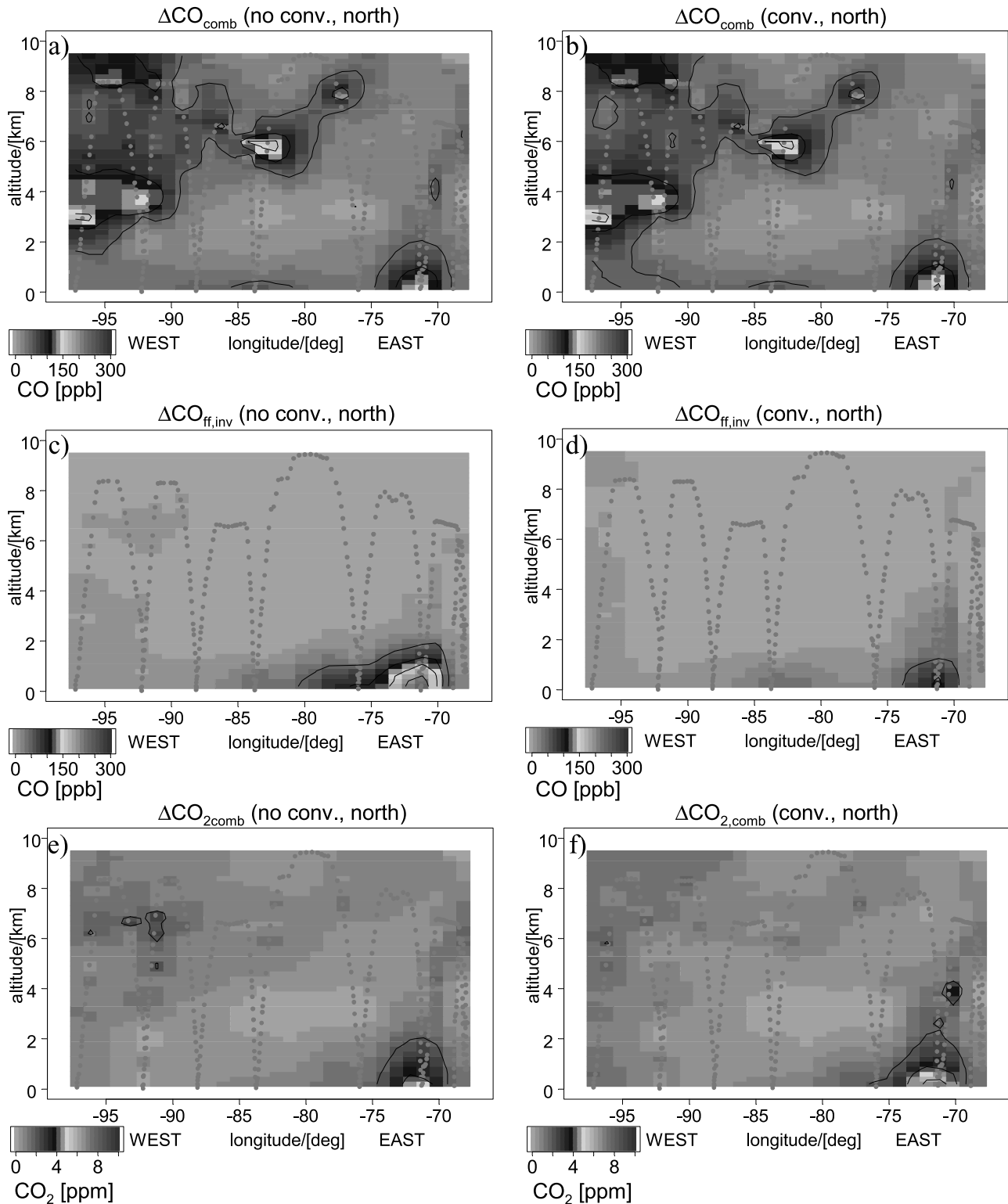
Figure 6. Combustion signals for CO and CO_2 along the northern survey. (a and b) CO fossil fuel signal ΔCO_{comb} after equation (11), (c and d) CO fossil fuel signal $\Delta CO_{ff,inv}$ calculated by ROAM using the emission inventory, and (e and f) CO_2 combustion signal after equation (10), using measured CO and CO_2:CO emission ratio. Left column (Figures 6a, 6c, and 6e) shows results for the non-convective case, right column (Figures 6b, 6d, and 6f) for the convective case. To interpolate the tracer data in altitude and longitude between the measurement locations, the squared inverse of the distance was used as weighting, with the distance measured in units of 500 m vertical and degree longitude horizontal (aspect ratio 1/200). Tracer data from different days and different times of the day were used for the cross sections. See color version of this figure at back of this issue.

increased the uncertainty by a factor of 2 to account for the fact that in these cases, inventory-based CO$_2$ fossil fuel signals ($\Delta CO_{2,ff,inv}$), inventory-based CO fossil fuel signals ($\Delta CO_{2,ff,inv}$) and the estimate of the biomass burning emission ratio were used to derive $\Delta CO_{2,comb}$. Uncertainties in $\Delta CO_{2,comb}$ calculated in this way range from 0.2 to 3 ppm, and for CO$_2$ fossil fuel signals larger than 5 ppm, the relative uncertainty is 35%.

[71] Figures 7a and 7b show the signal due to biosphere-atmosphere exchange ($\Delta CO_{2,veg}$), derived from the measured CO$_2$ distribution, combustion CO$_2$, and advected background using equation (8). The vegetation signal, on average, is an order of magnitude larger than the contribution from combustion. The strongest vegetation signals are found in the lower atmosphere of the northern survey, with a drawdown of 13–18 ppm in the west and 10–15 ppm in the eastern part. At the lowest altitudes over Massachusetts (71° W) a large, positive signal of about 10 ppm was observed during the takeoff of the aircraft in the morning (8 am local time), caused by accumulation of respiration in a shallow mixed layer during the previous night.

[72] Slightly negative values of $\Delta CO_{2,veg}$ are found in the free troposphere, with lowest values of −5 ppm observed in the middle part (84°W) of the northern survey. This appears to be a continental-scale signal of CO$_2$ uptake by vegetation, transported vertically to the middle troposphere by convection. The vegetation signal in the southern survey shows free tropospheric values close to zero, and a large buildup of CO$_2$ in the lowest kilometer, exceeding 10 ppm over Kansas. Significant drawdown in CO$_2$ along the southern survey was only observed at the northeastern terminus, over Massachusetts and Maine; vegetation signals at this time were roughly −10 ppm in both the lower atmosphere and the free troposphere, indicating convective transport.

[73] Uncertainties in $\Delta CO_{2,veg}$ (Figures 7c and 7d) are derived for each receptor location by propagating errors in equation (8) from the different terms, which are assumed to be independent. The uncertainty in the advected background field was calculated as a sum of the residuals (1.15 ppm, section 2.2.1) and an additional term, which accounts for the omission of the zonal longitudinal gradients. The latter term was assumed to increase at 1 ppm per 6000 km distance between the position of the particle ensemble at 15 days back and the location of the lateral boundary field at 145 W. The overall uncertainties in $\Delta CO_{2,veg}$ are much larger than the ±0.19 ppm from the measurement uncertainty for mixed-layer averaged CO$_2$ (see section 2.2). However, the uncertainties are significantly smaller than the vegetation signal itself (see Figure 7), with relative uncertainties approaching 5% for unpolluted areas with strong signature.

4.3. Upscaling Ameriflux From a Bayesian Inversion Analysis

[74] The derived vegetation signals $\Delta CO_{2,veg}$ are now used as the observational constraints for the scaling parameters in the biospheric flux model; that is, we use equations (15) and (16) to obtain optimal estimates and posterior uncertainties of the scaling parameters for the gross fluxes (GEE and R). The upscaling was done separately for the northern survey and the southern survey in order to account for the large differences in the vegetation condition index for the different footprint areas. The Northeastern part of the southern survey (east of the dashed line in Figure 7b) was included in the northern survey, since the footprints for these measurements were located further north than the rest of the southern survey (Figure 4). Relative influences from the different vegetation classes for mixed-layer measurements (Table 2) indicate that the dominant classes are the Northern Pacific Ocean and forests, followed by shrublands and croplands; less than 1% of the influence was from wetlands, which we therefore ignored in the following analyses. Relative contributions of the different vegetation classes to the total GEE and respiration signals (Table 2) based on prior scaling factors show that forests and croplands are the main contributors to uptake and release of CO$_2$ during this time of the year, accounting for about 90% of the signal.

[75] Uncertainties due to the different types of model error (S_{part}, S_{eddy}, S_{transp}, S_{aggr}, and S_{ocean}) calculated according to section 2.4 are shown in Figures 7e and 7f for the northern and southern survey. They are significantly larger then the uncertainty in the vegetation signal (Figures 7c and 7d). However, it is important to mention that the model error (with the exception of S_{part} and S_{eddy}) was assumed to be correlated within individual profiles, thus reducing the overall uncertainty as compared to an uncorrelated error. Together with the prior uncertainty, these errors serve as weights for the optimization.

[76] Figure 8 shows the retrieved scaling factors for forest, crop and shrubland GEE and R in the footprint areas of the two regions (northern and southern survey), obtained for the cases with and without convection. Also shown are the uncertainties associated with the retrievals (square-roots of the diagonal elements of S_λ). Observed fluxes for forests in the northern footprint are in rough agreement with simple upscaling of AmeriFlux sites (scaling factors close to 1), while southern forests are significantly less active (scaling factor for GEE less than one), reflecting stressed vegetation. The strong overestimation of crop fluxes is expected, since a single AmeriFlux site within a crop field was assumed to be representative for fluxes of both "cropland" and "cropland/natural vegetation mosaic" classes in the IGBP vegetation grid. Scaling factors calculated for the convective case are different from the non-convective case. This is due to dilution of surface signals to higher altitudes and due to a footprint that is confined to a smaller area as well as a shorter time-span (see Figure 5). The differences, although not statistically significant, are on the same order of magnitude as the posterior uncertainties, indicating that uncertainties associated with convection are as important as the other components of the measurement error S_ε.

[77] Significant reduction in the prior uncertainty was achieved for most of the scaling factors. Of course this depends on the prior uncertainty which was conservatively estimated. However, the reduction of the priori uncertainties also indicates the amount of information contained in continental CO$_2$ vegetation signals. Posterior uncertainties range from 0.2 for forest GEE, to more than 3 for shrubland GEE. The retrieved scaling factors for shrubland are below zero, indicating that GEE and R could not be separated properly for this vegetation class (corresponding off-diagonal elements of the S_λ are on the order of 0.9).

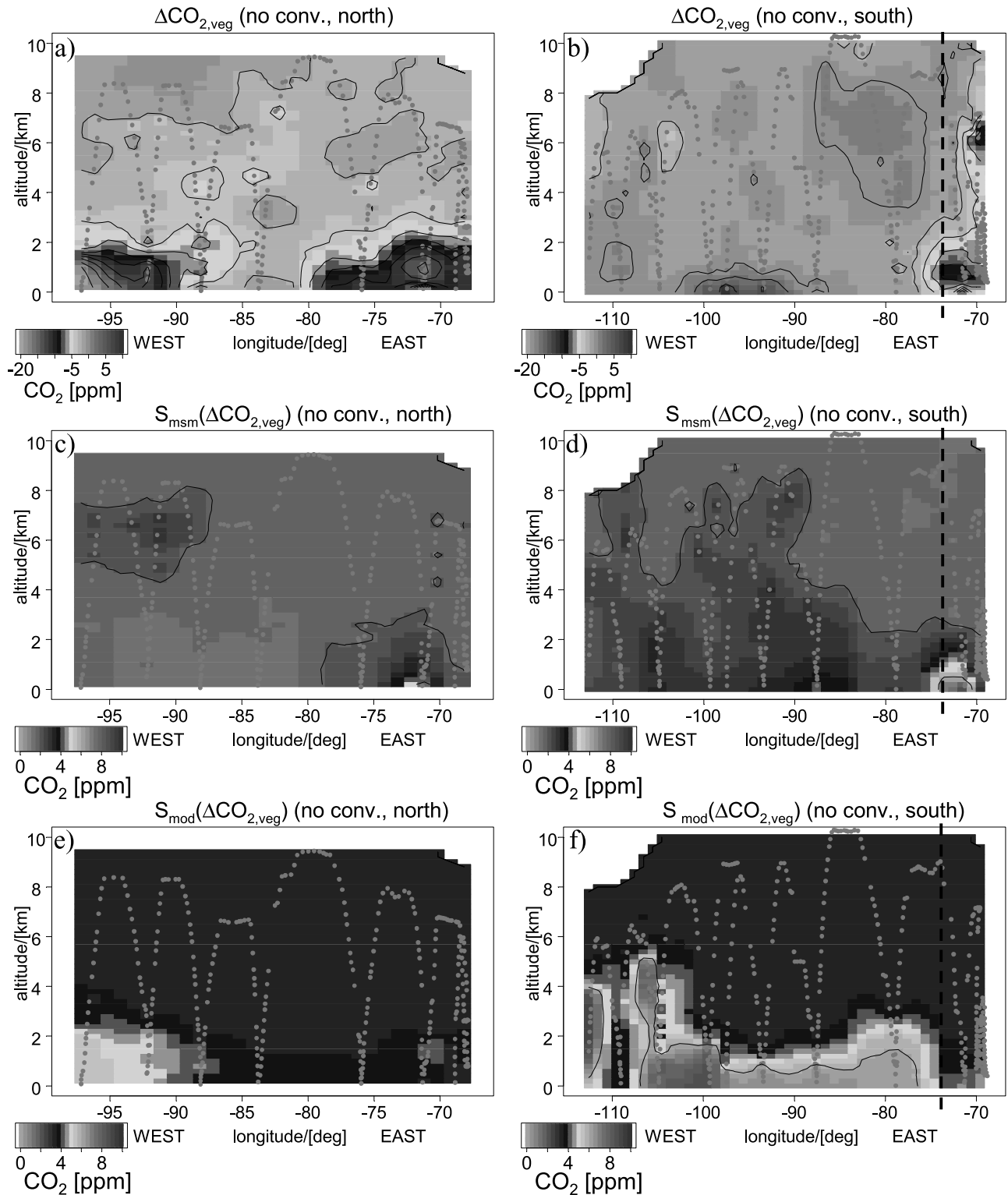[78] NEE for the whole surface domain was calculated on a 3-hourly time step for the month of August 2000 on the

Figure 7. Measurement derived CO_2 vegetation signals after equation (9), for the (a) northern survey and (b) southern survey, (c and d) uncertainty of the CO_2 vegetation signal, and (e and f) uncertainty of the modeled vegetation signal due to the model error components S_{part}, S_{eddy}, S_{transp}, S_{aggr}, and S_{ocean}. All results are shown for the non-convective case. The vertical dashed line in Figure 7b marks the eastern end of the data used in the optimization for the southern survey. See color version of this figure at back of this issue.

Table 2. Relative Influence From Different Vegetation Classes, for Runs Without Subgrid Convection[a]

Vegetation	Influence		GEE Signal		Respiration Signal	
	North Survey	South Survey	North Survey	South Survey	North Survey	South Survey
Water bodies	44%	38%	2%	1%	NA	NA
Forests	16%	21%	44%	71%	41%	61%
Shrublands	29%	29%	8%	3%	24%	10%
Croplands	11%	11%	45%	25%	34%	21%
Wetlands	<1%	<1%	NA	NA	NA	NA
Wetlands	<1%	<1%	NA	NA	NA	NA

[a]Also shown is the relative contribution of each vegetation class to the signal due to assimilation fluxes (GEE signal) and due to respiration fluxes (respiration signal).

basis of equation (11), using surface temperature and downward shortwave radiative flux from the EDAS fields and scaling factors from the non-convective case. The resulting estimates for the 24-hour averaged net biosphere-atmosphere fluxes for the non-convective case are shown in Figure 9. Diurnally averaged fluxes range from 6 μmole/m^2/s uptake to 2 μmole/m^2/s release of CO_2. Note that, in this illustration, the parameters derived from the northern survey were applied to the whole domain, for the later time period, with southern survey parameters used for the earlier time period.

[79] Error bars for individual grid elements are large, since in addition to the uncertainties for the retrieved scaling factors, the temporal and spatial upscaling from the footprints to the larger regions have to be taken into account. We therefore regard the maps in Figure 9 as hypothetical snapshots of NEE that are subject to large uncertainties but capture some of the spatial and temporal variability associated with the response to light and temperature.

[80] Most striking is the variability of fluxes in both space and time: in the footprint area of the northern survey (Figure 9, top) patches of low uptake are located next to areas with strong uptake, and these areas change from day to day. This variability is dominated by moving cloud systems: on 17 August a band of clouds covered the Great Lakes and areas north and moved to the east during the following days, reducing photosynthetic uptake. In the footprint area of the southern survey simulated fluxes are small, with less temporal variability. Spatial variability was associated with vegetation patterns in the western end of the flight area.

4.4. Representation Error Over the Continent

[81] In the companion paper [*Gerbig et al.*, 2003] a measurement-derived representation error was presented, based on the spatial variability of CO_2 at different horizontal scales (also shown in Figure 10). High variability of upstream surface fluxes on small spatial scales is a possible cause for this behavior. To assess the spatial variability in CO_2 mixing ratios associated with fine-scale variance of the surface fluxes, we ran the receptor-oriented model with different spatial resolutions (from 20 to 640 km). The model was run for the same mixed-layer locations that were used in the statistical analysis of the representation error, and mixing ratios representing a continental flux signal were calculated using our posterior biospheric fluxes and the emission inventory for CO_2 fossil fuel. For each receptor we calculated differences between these signals from the high-resolution run (20 km biosphere, 1 deg. fossil fuel) and from the degraded runs. Standard deviations of these differences

are shown in Figure 10 (black solid lines) for the different resolved grid scales of the degraded runs (note that the results are shifted to match the representation error at 20 km grid size). These standard deviations match the measurement-derived representation error at scales up to 160 km.
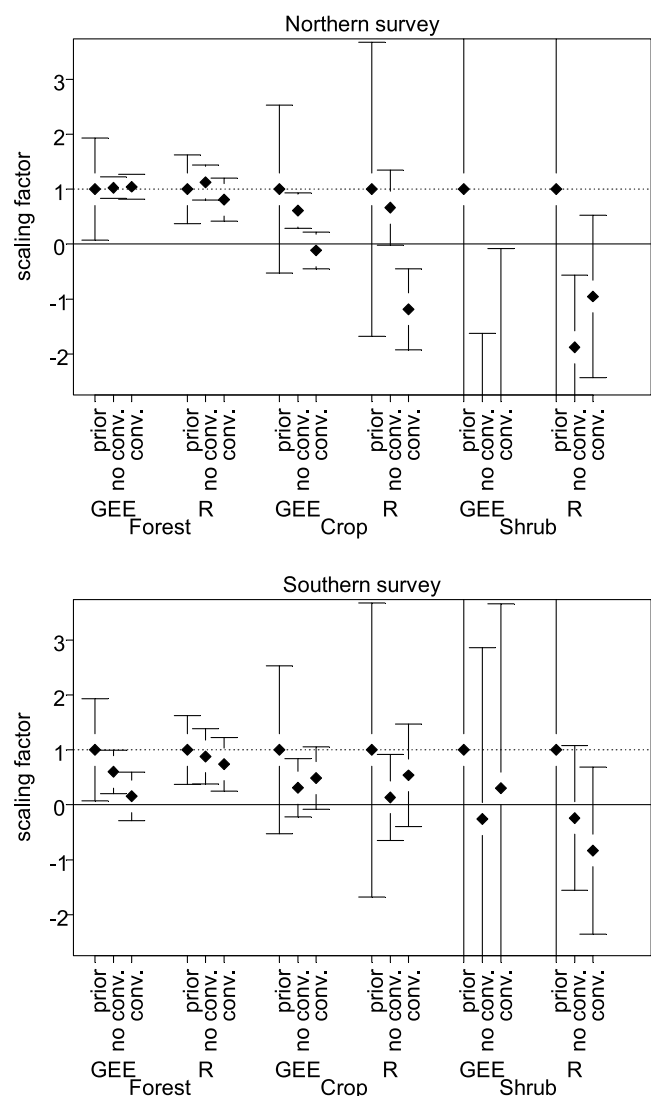

Figure 8. Inversion results from the Bayesian inversion applied to the COBRA data. Shown are the estimates of the scaling factor (diamonds) and associated uncertainties (error bars).
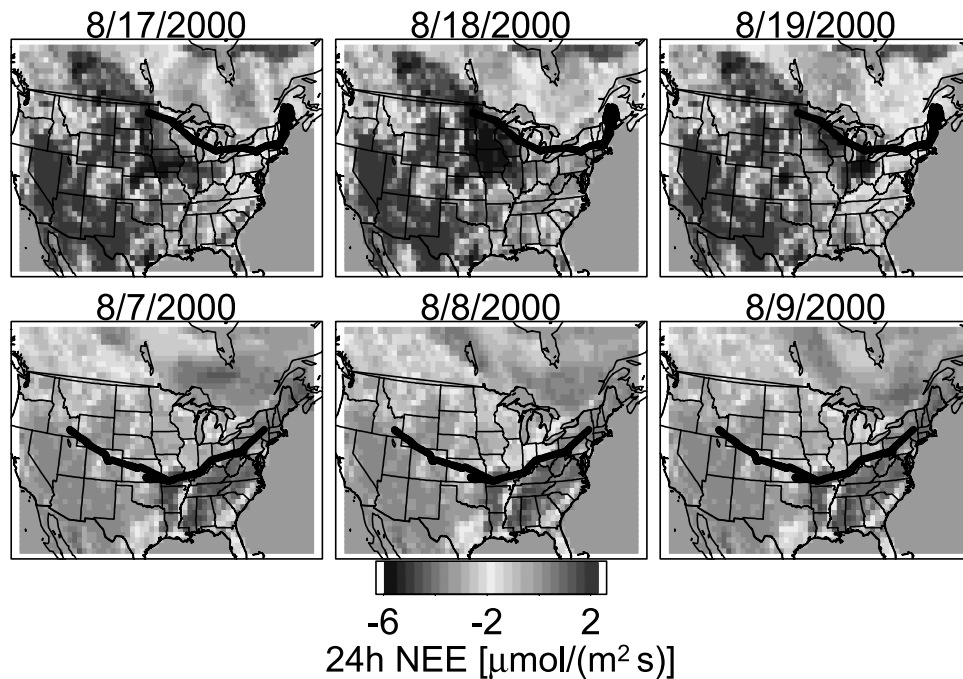
Figure 9. Twenty-four-hour averages of NEE from forests and croplands based on the upscaling of AmeriFlux data for three consecutive days preceding the measurements (results for non-convection case). The top and bottom rows show results using scaling factors constrained by the northern and the southern surveys, respectively. See color version of this figure at back of this issue.

We infer that the increase in the representation error of CO_2 in the scale range from 20 km to 160 km is due to spatial variability of upstream surface fluxes.

[82] We also calculated these differences for runs with shortened integration time, accounting only for surface fluxes 6 and 1 hours prior to the measurement (shown in Figure 10 as the dotted and dashed-dotted lines, respectively). The results for 6 hour integration time closely follow the 360 hour results at small scales up to 80 km, suggesting that most of the variability in CO_2 on these scales is due to variability of fluxes in the near field of the observations. However, results for 1 hour integration time tend to deviate from the results for longer integration times at all scales, starting at 40 km.

[83] Differences between the high-resolution run and a run which only resolves surface fluxes at a coarse scale can also be interpreted as an estimate for the aggregation error: signals from surface fluxes with no spatial variation on the coarse scale are compared with signals from fluxes with resolved spatial variations on the 20 km scale. For a coarse scale of 2000 km (corresponding to the size of the footprints) the standard deviations of differences are about 3 ppm. This value provides an estimate of the aggregation error.

5. Discussion

[84] Many attempts have recently been made to derive regional CO_2 fluxes from long-term tower based measurements of CO_2 in simpler budget approaches. These approaches usually use a one-dimensional framework and combine a measure of the turnover time of the mixed layer with the free troposphere with temporal variations in differ-
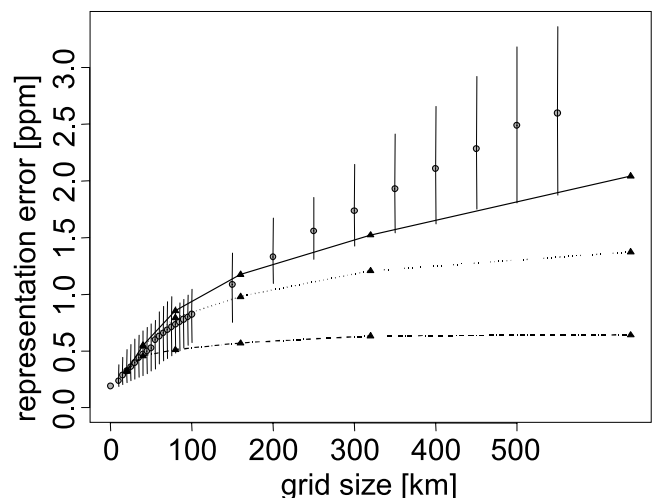

Figure 10. ROAM generated representation error for mixed-layer receptors (triangles) as a function of surface flux grid size. Results are shown for integration times of 360 hours (solid line), 6 hours (dotted line), and 1 hour (dashed-dotted line). Also shown is the measurement derived total representation error (combined measurement uncertainty and representation error) after *Gerbig et al.* [2003] (light grey circles and vertical bars indicate the 5–95% range). Values for the ROAM generated representation error are offset to match the measurement derived representation error at a grid size of 20 km.

ences between mixed-layer CO_2 and free tropospheric CO_2 to estimate fluxes over an unspecified area. The receptor-oriented analysis of the COBRA data indicates a number of problems associated with such simplifications: (1) CO_2 in the free troposphere over the continent cannot be described simply by the MBL-reference CO_2 taken at the same latitude, since significant biases can result from both vertical gradients and meridional components in advection (see section 2.2; also the companion paper); a more sophisticated boundary condition is needed to reduce these biases; (2) Footprints for mixed-layer receptors typically extend over an area of about 10^5 km^2 (Figure 5) and over a few days preceding the measurement (Figure 6), and changes in these footprints due to meteorological conditions will cause changes in CO_2 due to spatial variability in surface fluxes; these variations in CO_2 need to be properly attributed to the areas sampled, rather than to changes in fluxes in a single area [*Lin et al.*, 2003]; (3) local estimates of a turnover time between mixed layer and free troposphere at the receptor cannot be assured to be equivalent to turnover times in the footprint area. Hence we believe that an analysis framework that appropriately represents atmospheric transport and spatially and temporally varying surface fluxes should be used to infer regional fluxes.

[85] The measurement-derived vegetation signal ($\Delta CO_{2,veg}$) presented in section 4.2 from the COBRA in-situ measurements of CO_2 and CO provides a tight constraint for regional biospheric fluxes, with uncertainties on the order of 15% for strong mixed-layer signals in absence of pollution (Figures 7a–7d). To assess how successfully the approach resolves the spatial and temporal variability in CO_2 mixing ratios, we compare simulated CO_2 vegetation signals $\Delta CO_{2,veg,mod}$ with "measured" vegetation signals $\Delta CO_{2,veg}$. The simulated distribution of CO_2 vegetation signals $CO_{2,veg,mod}$ for the different surveys and the different model cases are shown in Figures 11a–11d. In general there is reasonable agreement between the modeled (Figure 11) and the "measured" (Figure 7) cross sections. Both exhibit the strong depletion of CO_2 in the lower atmosphere for the northern survey, and a build up of CO_2 in the middle of the southern survey. $CO_{2,veg,mod}$ explained 40% of the variance in $CO_{2,veg}$. However, there are numerous discrepancies which can be related to (1) inadequate representation of atmospheric transport and (2) oversimplified representation of biosphere. In the following paragraphs we discuss errors in modeled vegetation signals due to points 1 and 2 and their effects on the simulated large-scale distribution of vegetation CO_2.

5.1. Discrepancies Related to Transport Representation

[86] Transport errors can arise from inadequate parameterizations within STILT (e.g., convection, mixed-layer height, turbulence statistics) and from errors in the assimilated winds used to drive STILT, either because of the assimilation process itself or by inadequate post-processing [*Lin et al.*, 2003]. The only transport uncertainty accounted for in the inversion is the error due to the mismatch in z_i, which we attempted to represent in S_{transp} (section 4.3). The other errors are more difficult to account for but are expected to cause discrepancies between modeled and measured vegetation signals.

[87] The clearest discrepancies are caused by the extremes in vertical redistribution built into the convective parameterization: in the westernmost profile of the northern transect the model case with convection (Figure 11c) overestimates the vertical dilution of $\Delta CO_{2,veg,mod}$, while the non-convective case (Figure 11a) shows reasonable agreement with the observations (Figure 7a). For the easternmost profile of the southern transect it is the opposite: no vegetation signal was found around 6 km in the non-convective case, while a significant depletion of CO_2 resulted from the convective case (see Figures 7b, 11b, and 11d). In this example convection was active in the footprint area during the afternoon on the day before the flight. This indicates that the two cases (excessive convection and no convection) bound the true influence of convection on the tracer distribution. However, the strong impact of convection on the derived fluxes (Figure 9) indicates the necessity for a convection scheme to accurately quantify convective cloud transport, with realistic temporal variations to avoid biases in simulated CO_2 due to covariation with surface fluxes. Clearly assimilated meteorological products should include convective fluxes, and the convective parameterizations in the assimilation model need to be realistic.

[88] In the middle of the northern survey (at 83 W) neither the nearly flat profile of $\Delta CO_{2,veg}$ with a depletion in the lowest 2 km of only 2–3 ppm nor the free tropospheric vegetation signals of −1.2 ppm is properly reproduced by either model case. Both seem to produce a minimum depletion in the mixed layer similar to the observations, but displaced hundreds of km further to the west, at 88 W. A potential explanation for this discrepancy can be seen from Figure 9: surface fluxes are strongly influenced by a cloud system moving from the Great Lakes region toward the northeast, and the amount of CO_2-depletion in the sampled air depends critically on the relative location of footprint area and the cloud system at a given time. Reduced cloud cover (or higher incident radiation) over the footprint area in the simulation can cause substantial changes in CO_2 vegetation signals. A further contribution to this discrepancy might be unresolved meso-scale circulations in the proximity of the Great Lakes, yielding wrong estimates for the footprint close to the receptors, where influence is strongest (see Figure 4). Significant errors can accrue if the assimilation model misrepresents clouds in the footprint, as might be expected around the Great Lakes, or in areas near frontal systems.

[89] Throughout the profiles of both surveys there is a significant underestimation of the nighttime build-up of respiration signal in the early morning mixed layer, and an overestimation at levels in the residual layer. The problem results from slightly too much vertical mixing: residual layer receptors at altitudes of around 500 m agl are influenced by the surface during night time in the dispersion model, whereas the strong vertical gradients of tracer indicate a stronger mixing barrier at low levels in the atmosphere. The ability to use the information contained in the nighttime buildup of respired CO_2 is crucial to independently constrain fluxes due to nighttime release and daytime uptake of CO_2. The difficulty in partitioning respiration and uptake of CO_2 is illustrated by the large correlation (r ~ 0.9) between scaling factors for uptake (GEE) and respiration (R). A better separation between
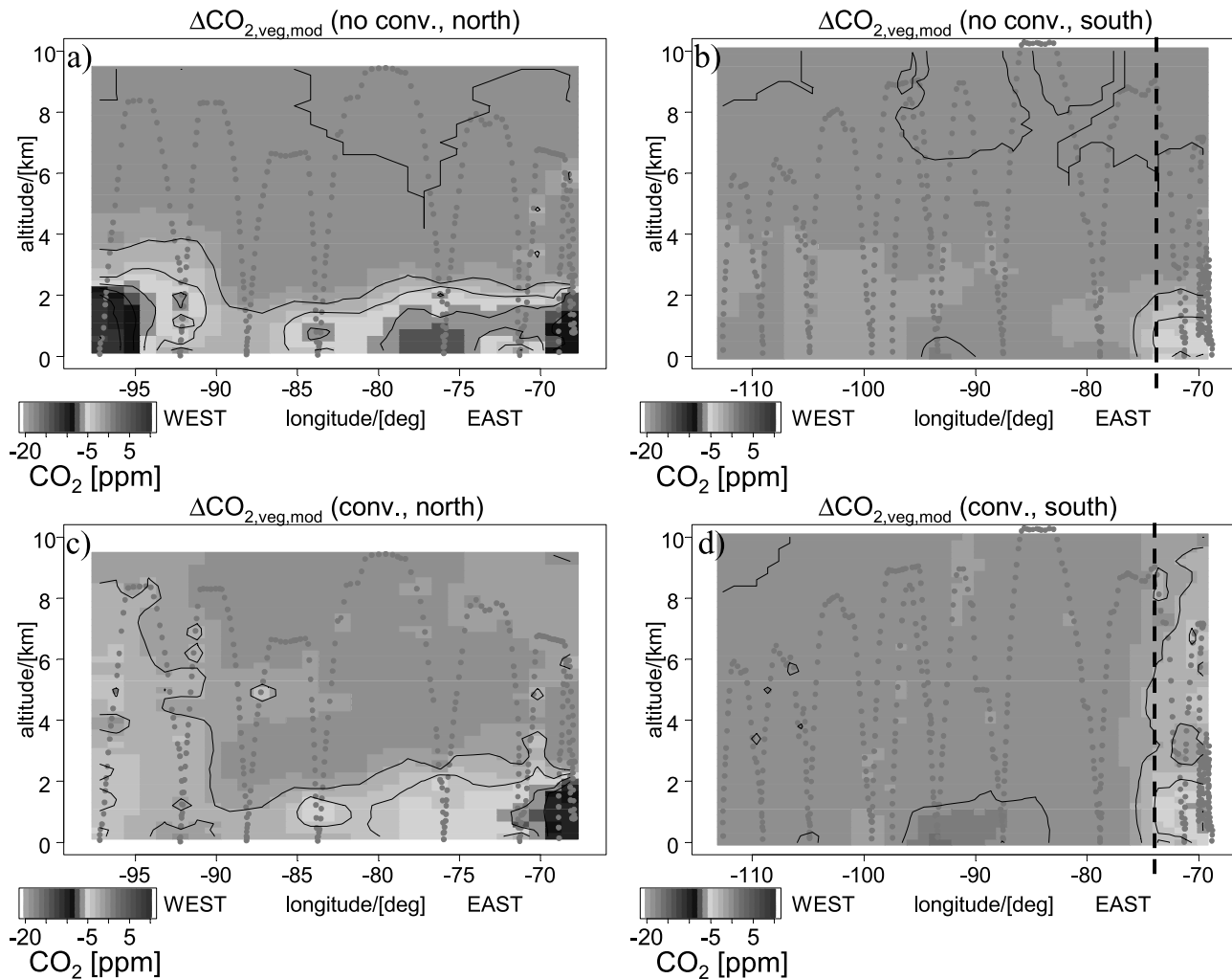
Figure 11. Modeled vegetation CO2 signal (a and b) for the non-convective case and (c and d) for the convective case. Left column (Figures 11a and 11c) shows results for the northern survey, right column (Figures 11b and 11d) for the southern survey. Left of the vertical dashed lines in Figures 11b and 11d the southern survey scaling factors are used; to the right of these lines, northern survey scaling factors were applied. See color version of this figure at back of this issue.

nighttime release and daytime uptake can only be expected from more sophisticated mixing schemes in models during night time, and from improved vertical resolution.

[90] As discussed by *Lin et al.* [2003], lack of mass conservation in EDAS winds leads to inaccuracies in source-receptor relationships. These inconsistencies are related to the post-processing of wind fields (degradation of resolution and non-mass conserving transformations in the vertical coordinate and horizontal re-projections), and to the assimilation process in the weather prediction model itself. Violations of mass conservation are likely to have an impact on the retrieved vegetation CO$_2$ signals, but the magnitude of the error is difficult to estimate. Using mass conserving wind fields, or correcting wind fields to conserve mass has therefore a high priority.

5.2. Discrepancies Related to Representation of the Biosphere

[91] The assumption made in the simple parameterization of biosphere-atmosphere flux is that a given vegetation class

within the footprint area for a given survey (north or south) has a light and temperature response similar to the Ameri-Flux data from July and August 2000 representing this vegetation class. The assumption of spatial homogeneity within a vegetation class over the footprint areas re-introduces an aggregation error, problematic given significant spatial differences in vegetation health, and the broad range of vegetation types included in one of our coarse classes, within a footprint (Figure 2e of *Gerbig et al.* [2003] and Figure 5). The estimated uncertainty of 3 ppm due to aggregation (section 4.4) might account for the aggregation error, but it significantly increases the model error by on average 70%.

[92] There are two basic approaches to reduce this source of aggregation error: one can decrease the size of the flux region, and thus solve for more parameters which can adjust the spatial flux pattern, or one can improve the spatial distribution of fluxes, e.g., by adding information on vegetation health, or by using a full biosphere model which describes phenomena such as drought stress. The former

approach of course can lead to an under-determined problem; in this case it would be required to add information about spatial and temporal correlations of the prior uncertainties.

[93] Future work will include the implementation of a more sophisticated representation of the biosphere, e.g., using information on leaf nitrogen, on crop data from USDA (e.g., harvesting cycles), and perhaps the use of satellite-derived quantities such as the vegetation condition index [Kogan, 1997] and NDVI (normalized difference vegetation index). The important capability of this receptor-oriented framework, however, is that it can be coupled to any surface flux patterns at high spatial and temporal resolution, such as that generated by a full biospheric model. Ultimately, one would not optimize for fluxes on some regional scale, but for parameters controlling uptake and respiration in a spatially explicit biosphere model.

6. Concluding Remarks

[94] COBRA can be viewed as a snapshot of data that could be provided by future sampling networks with enhanced spatial coverage and multiple profiles. We have analyzed COBRA data to critically test the requirements of a model-data fusion system aimed at deriving regional to continental flux estimates from measurements of CO_2 over the continent. The spatial patterns in the CO_2 data collected over the United States during COBRA show clear signatures of terrestrial fluxes during the active growing season. A significant fraction of the information in these signatures, however, is contained in relatively small spatial and temporal scales. Appropriate analysis frameworks to resolve these scales are needed in order to effectively use the information in continental measurements of CO_2; otherwise this fine-scale signal is transformed into large-amplitude unresolved variance.

[95] The receptor-oriented framework presented here was designed as an attempt to represent the processes that could potentially introduce biases in flux estimates and to extract the signal of terrestrial fluxes from continental CO_2 observations. It consisted of (1) a Lagrangian transport model (STILT) to couple transport and surface fluxes at high resolution using assimilated winds, (2) lateral boundary conditions for CO_2 and CO from a climatological analysis, and (3) surface fluxes from fossil fuel and the biosphere. This framework is analogous to a regional adjoint model in a Eulerian framework; both require boundary conditions from either a global model or from a climatology based on measurements. One difference is that the receptor-oriented approach resolves variations in surface flux on scales smaller than the grid of the meteorological fields.

[96] The receptor-oriented framework allowed us to calculate vegetation signals $\Delta CO_{2,veg}$ along the COBRA flight track as the difference between measured CO_2 and the sum of advected boundary mixing ratios and combustion signal. Simultaneous observations of CO enabled improved estimates of the combustion signal $\Delta CO_{2,comb}$. The vegetation signals provided a tight constraint for large-scale biosphere-atmosphere exchange fluxes from upscaling of AmeriFlux measurements. Modeled vegetation signals explained 40% of the observed variance, reasonable agreement in light of the shortcomings in current transport fields. The optimized

biosphere model provided estimates of net fluxes at continental scale, constrained to be consistent (subject to the limitations in the current analysis framework) with observed concentrations over the continent and with AmeriFlux data on vegetation sources and sinks.

[97] Discrepancies between modeled vegetation signals and vegetation signals derived from airborne observations revealed a variety of problems which have to be addressed in future studies: (1) an accurate convection scheme is required for a long-lived tracer like CO_2 with surface sources/sinks, whose atmospheric signatures are sensitive to vertical redistribution; (2) improvements in mixing schemes during nighttime periods (including higher vertical resolution) are needed in order to separate respiratory release and photosynthetic uptake of CO_2; (3) offline transport models using assimilated winds can be subject to violations of mass conservation, which can have a significant effect on the accuracy of source-receptor relationships; (4) large variations in biosphere-atmosphere exchange result from the reduction of downward radiation due to cloud systems, and it is important to accurately represent these systems and their effect on radiation in the footprint area of CO_2 measurements; (5) the overly simplified biospheric representation, assuming spatially and temporally constant light and temperature responses, is inadequate to yield accurate relationships between large-scale and local fluxes. Fluxes from a full biosphere model can and should be coupled to the receptor-oriented framework in the future, with focus on estimating biospheric parameters rather than scaling factors in the values of fluxes.

[98] When examining these large-scale measurements as a snapshot of the data that would be collected by a future sampling network, one would be interested in quantifying not only large fluxes during the growing season, but more importantly long-term (e.g., net annual) exchange fluxes over regional to continental scales. Thus the uncertainties have to be put in a long-term context, and it becomes very important how the errors are correlated from day to day, i.e., whether the uncertainties accumulate to a bias on seasonal to annual timescales, or whether the uncertainties are only correlated over short timescales, such that integration over longer time periods reduces the resulting uncertainties. One cannot assume that it is appropriate to average over unresolved spatial and temporal variability, because significant biases may be introduced: e.g., coarse vertical resolution and inappropriate representation of vertical mixing may cause biases because of the rectifier effect [Denning et al., 1996]. Biases can also result when systematic patterns in wind direction are not properly resolved because of anisotropy of CO_2 fluxes in the near field surrounding a measurement station. Therefore an appropriate description of the spatial and temporal covariances of the errors in addition to their magnitudes is required in the future.

[99] For the transport error, for example, this could be done involving a careful statistical analysis of differences between analyzed meteorological fields and extensive meteorological measurements, similar to what is done in weather prediction centers. A similar effort is required to describe prior uncertainties and their covariances in biospheric flux models and in emission inventories (which should also be included as a "soft" constraint [Engelen et al., 2002; Rodgers, 2000], e.g., by solving for emission

fluxes given a prior estimate). Future programs such as the NACP will provide data with enhanced temporal and spatial coverage which will allow us to assess these unresolved issues.

[100] We envision a variety of future applications for the receptor-oriented framework presented here. The framework can be applied to long-term measurements (e.g., from instrumented towers) of CO_2, CO, and other tracers like CH_4 to provide information on regional fluxes over annual and interannual timescales. The capability to calculate quantitative footprint estimates, resolving influences from different vegetation types, is invaluable to evaluate the design of hypothetical observational networks. When coupled to a full biospheric model, the framework could enable assimilation of atmospheric observations to constrain parameters controlling both short-term, rapid processes and slow processes in the carbon cycle at continental and regional scales.

Appendix A

A1. Fitting of Upstream CO_2 Data Over the Pacific

[101] The data were grouped into two latitude bands (south and north of 37.5°N), and into altitude bins, with ranges 0–2.5, 2.5–5, 5–8, and above 8 km. To propagate the CO_2 time series vertically, daily averaged mixing ratios $CO_{2,z}(t, lat)$ measured within altitude bin j and at latitude bin i were related to the time-series at the surface $CO_{2,S}(t', lat)$ by the Green's-function $G_{i,j}(t - t')$ for the 1-D advection-diffusion equation, similar to the approach used by *Andrews et al.* [1999] for stratospheric CO_2:

$$CO_{2,z}(t, lat) \int G_{i,j}(t - t') \cdot CO_{2,s}(t', lat) \cdot dt \quad \text{(A1)}$$

with:

$$G_{i,j}(t - t') = \frac{a}{\sqrt{(t - t')^3}} \exp\left(-b_{i,j}(t) \cdot (t - t') - \frac{c_{i,j}(t)}{t - t'}\right) \quad \text{(A2)}$$

[102] Parameter $b_{i,j}(t)$ controls the decaying behavior at large $t - t'$, while $c_{i,j}(t)$ controls the behavior at small $t - t'$. a is an integration constant to normalize G. The timescales $1/b_{i,j}(t)$ and $c_{i,j}(t)$ are allowed to be time dependent, parameterized by a sinusoidal with a one year cycle:

$$b_{i,j}(t) = \frac{b_{i,j}}{1 + s_i \cdot \sin(2\pi \cdot (t - t_0)/T)} \quad \text{(A3)}$$
$$c_{i,j}(t) = c_{i,j} \cdot (1 + s_i \cdot \sin(2\pi \cdot (t - t_0)/T))$$

[103] Here T is one year, and t_0 was empirically chosen as 1 October to allow for smaller timescales in summer (faster vertical transport). The parameter s_i controls the amplitude of the seasonal variation of the timescales. The parameters $b_{j,i}$, $c_{j,i}$, and s_i for the 2 latitude bands and the 4 altitude bins, i.e., a total of 18 parameters, were estimated by least square optimization using a total of 16,788 observations (see Table 2).

[104] This framework of a one-dimensional system with a boundary condition at the ground ($z = 0$) is not very realistic

(the sources for CO_2 at higher altitudes are usually not exactly underneath at the same horizontal location), but it is practical: It provides a way to relate upper air measurements of CO_2 to surface measurements with a certain time-lag (causing the seasonal cycle to occur later or earlier at higher altitudes), and with a damping of higher-frequency variations via the width of $G(t-t')$ which reflects the effect of atmospheric mixing processes. Further it makes use of the fact that CO_2 is conserved in the troposphere.

A2. Fitting of CO Data

[105] To propagate the CO time series vertically, a different approach had to be used, since CO (unlike CO_2) is not conserved in the troposphere. The data were grouped into three different latitude bands (10–37.5°N, 37.5–62.5°N, 62.5–70°N); these bands are centered over the surface stations. The CO time series at the surface $CO_S(t, lat)$ was separated into a slow component $CO_{slow}(t, lat_i)$ (timescales longer than 1 year) and a fast component $CO_{fast}(t, lat)$ (timescales 1 year to 40 days). The following parameterization was chosen to fit measured mixing ratios $CO_z(t, lat)$ at higher altitudes:

$$CO(t, z, lat) = CO_{slow}(t - f_i \cdot z, lat) + d_i \cdot z$$
$$+ (1 + e_i \cdot z) \cdot CO_{fast}(t - f_i \cdot z, lat) \quad \text{(A4)}$$

[106] Here, parameter d_i represents an average vertical CO gradient for latitude bin i, which is required to account for photochemical losses of CO, e_i reflects the damping of the seasonal cycle (and higher frequencies) with increasing altitude, and f_i is the change of the lag time with altitude. These 3 parameters for each latitude band were estimated by using a least square fit at each latitude bin from a total of 15,797 observations.

References

Andres, R. J., G. Marland, I. Fung, and E. Matthews, A one degree by one degree distribution of carbon dioxide emissions from fossil-fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cycles*, *10*(3), 419–429, 1996.

Andrews, A. E., K. A. Boering, B. C. Daube, S. C. Wofsy, E. J. Hintsa, E. M. Weinstock, and T. P. Bui, Empirical age spectra for the lower tropical stratosphere from in situ observations of CO_2: Implications for stratospheric transport, *J. Geophys. Res.*, *104*(D21), 26,581–26,595, 1999.

Atkinson, R., D. L. Baulch, R. A. Cox, F. R. J. Hampson, J. A. Kerr, M. J. Rossi, and J. Troe, Evaluated kinetic and photochemical data for atmospheric chemistry: Supplement V, *J. Phys. Chem. Ref. Data*, *26*, 521–1011, 1997.

Baldocchi, D., et al., 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*(11), 2415–2434, 2001.

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

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, Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.*, 101(D22), 29,239–29,253, 1996.

Denning, A. S., D. A. Randall, G. J. Collatz, and P. J. Sellers, Simulations of terrestrial carbon metabolism and atmospheric CO$_2$ in a general circulation model. Part 2: Simulated CO$_2$ concentrations, *Tellus, Ser. B*, 48, 543–567, 1996.

Ebel, A., R. Friedrich, and H. Rodhe (Eds.), *Transport and Chemical Transformation of Pollutants in the Troposphere*, vol. 7, *Tropospheric Modelling and Emission Estimation*, Springer-Verlag, New York, 1997.

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

Environmental Protection Agency (EPA), *Regional Interim Emission Inventories (1987–1991)*, vols. 1 and 2, Research Triangle Park, N. C., 1993.

Environmental Protection Agency (EPA), Inventory of U. S. greenhouse gas emissions and sinks: 1990–1999, Research Triangle Park, N. C., 2001.

Errico, R. M., What is an adjoint model?, *Bull. Am. Meteorol. Soc.*, 78(11), 2577–2591, 1997.

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

GLOBALVIEW-CO$_2$, *GLOBALVIEW-CO$_2$: Cooperative Atmospheric Data Integration Project: Carbon Dioxide* [CD-ROAM], NOAA Clim. Monit. and Diagnostics Lab., Boulder, Colo., 2002.

Holzer, M., and T. M. Hall, Transit-time and tracer-age distributions in geophysical flows, *J. Atmos. Sci.*, 57(21), 3539–3558, 2000.

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

Kogan, F. N., Global drought watch from space, *Bull. Am. Meteorol. Soc.*, 78(4), 621–636, 1997.

Kondo, H., N. Saigusa, S. Murayama, S. Yamamoto, and A. Kannari, A numerical simulation of the daily variation of CO$_2$ in the central part of Japan, *J. Meteorol. Soc. Jpn.*, 79(1), 11–21, 2001.

Law, R. M., et al., Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO$_2$ inversions, *Global Biogeochem. Cycles*, 10(4), 783–796, 1996.

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

Lefevre, N., A. J. Watson, D. J. Cooper, R. F. Weiss, T. Takahashi, and S. C. Sutherland, Assessing the seasonality of the oceanic sink for CO$_2$ in the Northern Hemisphere, *Global Biogeochem. Cycles*, 13(2), 273–286, 1999.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger, 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, 2003.

Marland, G., T. A. Boden, A. L. Brenkert, R. J. Andres, and J. G. J. Olivier, CO$_2$ from fossil fuel burning: Updates on the magnitude, distribution, and uncertainty of emission estimates, in *Fifth International Carbon Dioxide Conference*, p. 4, CSIRO, Cairns, Queensl., Australia, 1997.

Novelli, P. C., K. A. Masarie, and P. M. Lang, Distributions and recent changes of carbon monoxide in the lower troposphere, *J. Geophys. Res.*, 103(D15), 19,015–19,033, 1998.

Potosnak, M. J., S. C. Wofsy, A. S. Denning, T. J. Conway, P. C. Novelli, and D. H. Barnes, Influence of biotic exchange and combustion sources on atmospheric CO$_2$ concentrations in New England from observations at a forest flux tower, *J. Geophys. Res.*, 104(D8), 9561–9569, 1999.

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

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

Wang, Y. P., Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach: II. The atmospheric constraint, *Tellus, Ser.*, 290–304, 2003.

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

————————————

A. E. Andrews, NASA Goddard Space Flight Center, Mail Stop 916.0, Greenbelt, MD 20771, USA. (andrews@maia.gsfc.nasa.gov)

C. Gerbig, J. C. Lin, and S. C. Wofsy, Department of Earth and Planetary Sciences, Harvard University, Pierce Hall, 20 Oxford Street, Cambridge, MA 02138, USA. (chg@io.harvard.edu; jcl@io.harvard.edu; steven_wofsy@harvard.edu)

P. S. Bakwin and B. C. Daube, Climate Monitoring and Diagnostic Laboratory, National Oceanic and Atmospheric Administration, R/E/CG1, 325 Broadway, Boulder, CO 80305, USA. (pbakwin@cmdl.noaa.gov; bcd@io.harvard.edu)

C. A. Grainger, Department of Atmospheric Sciences, University of North Dakota, Box 9006, Grand Forks, ND 58202-9006, USA. (grainger@aero.und.edu)

B. B. Stephens, Atmospheric Technology Division, National Center for Atmospheric Research, Boulder, CO 80307, USA. (stephens@ucar.edu)
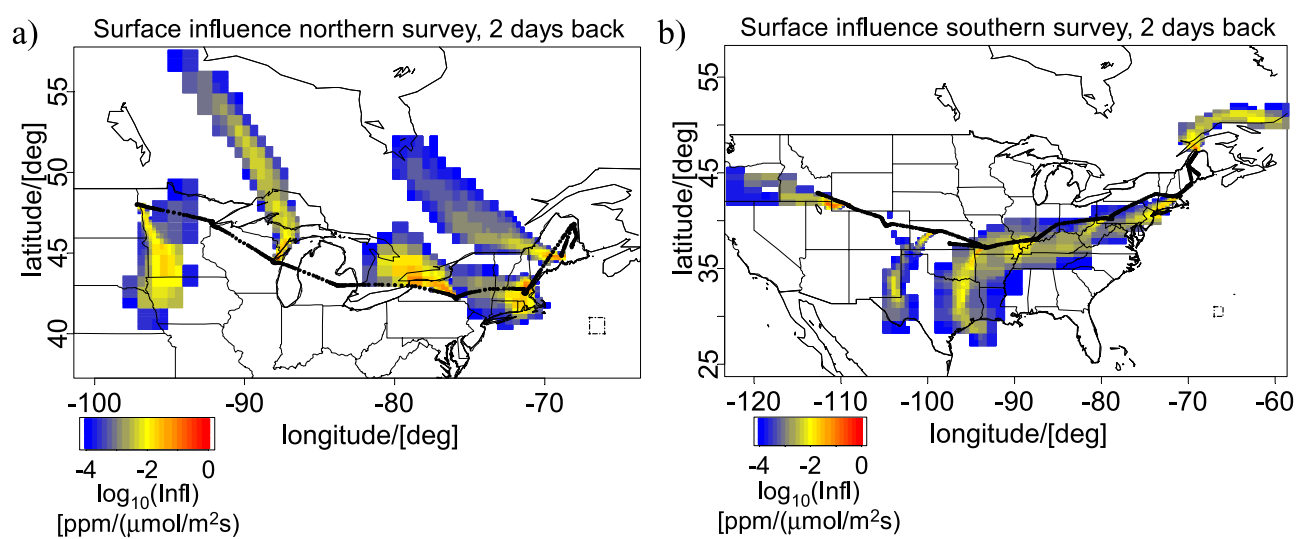
Figure 4. Footprints for mixed-layer receptors, integrated over the last 2 days prior to the measurements, calculated for the non-convective case for the (a) northern and (b) southern surveys. The different pixel sizes are a result of the dynamic grid resolution. The square with an edge length of 1 degree over the Atlantic Ocean is shown as reference.
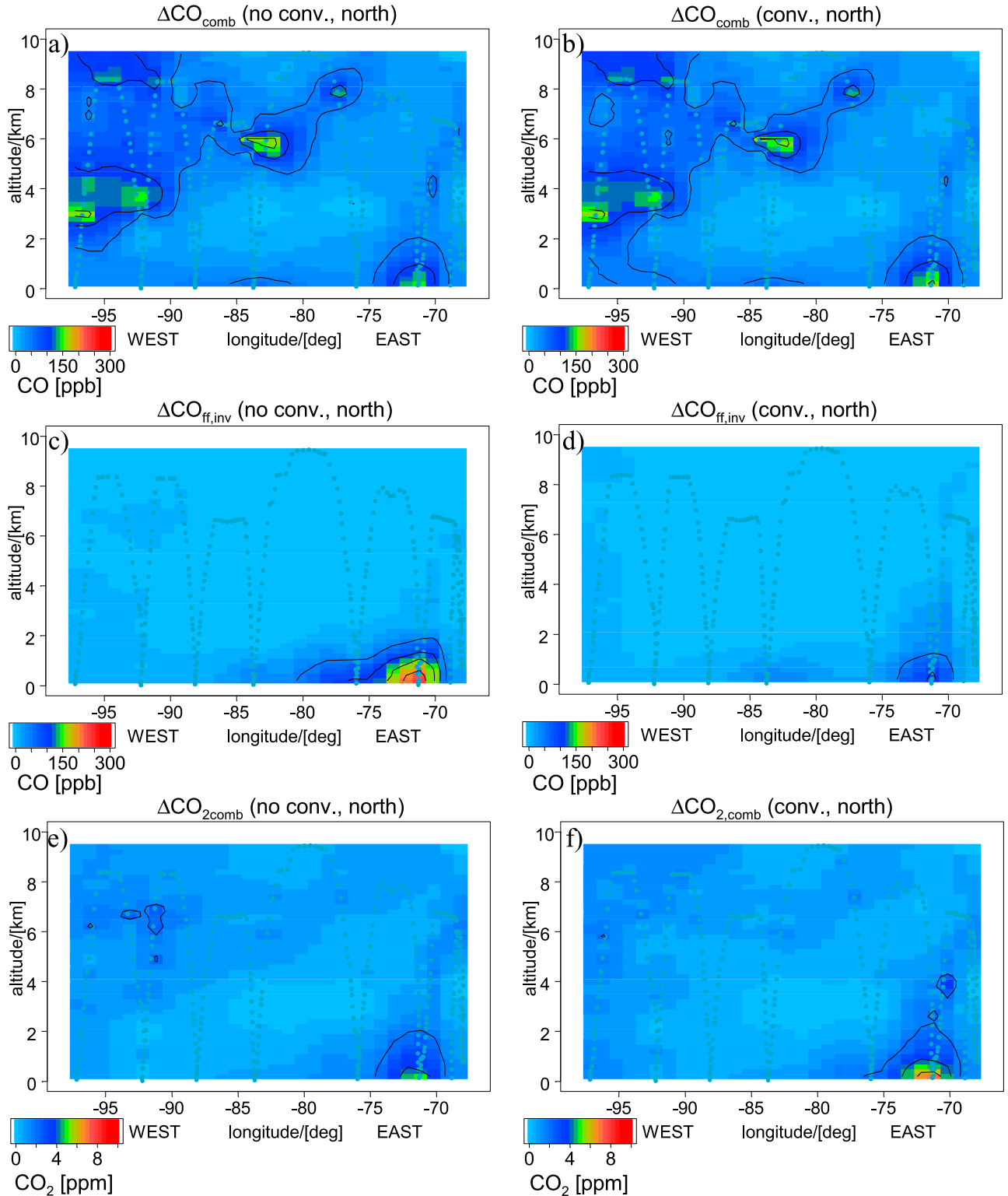
Figure 6. Combustion signals for CO and CO$_2$ along the northern survey. (a and b) CO fossil fuel signal ΔCO_{comb} after equation (11), (c and d) CO fossil fuel signal $\Delta CO_{ff,inv}$ calculated by ROAM using the emission inventory, and (e and f) CO$_2$ combustion signal after equation (10), using measured CO and CO$_2$:CO emission ratio. Left column (Figures 6a, 6c, and 6e) shows results for the non-convective case, right column (Figures 6b, 6d, and 6f) for the convective case. To interpolate the tracer data in altitude and longitude between the measurement locations, the squared inverse of the distance was used as weighting, with the distance measured in units of 500 m vertical and degree longitude horizontal (aspect ratio 1/200). Tracer data from different days and different times of the day were used for the cross sections.
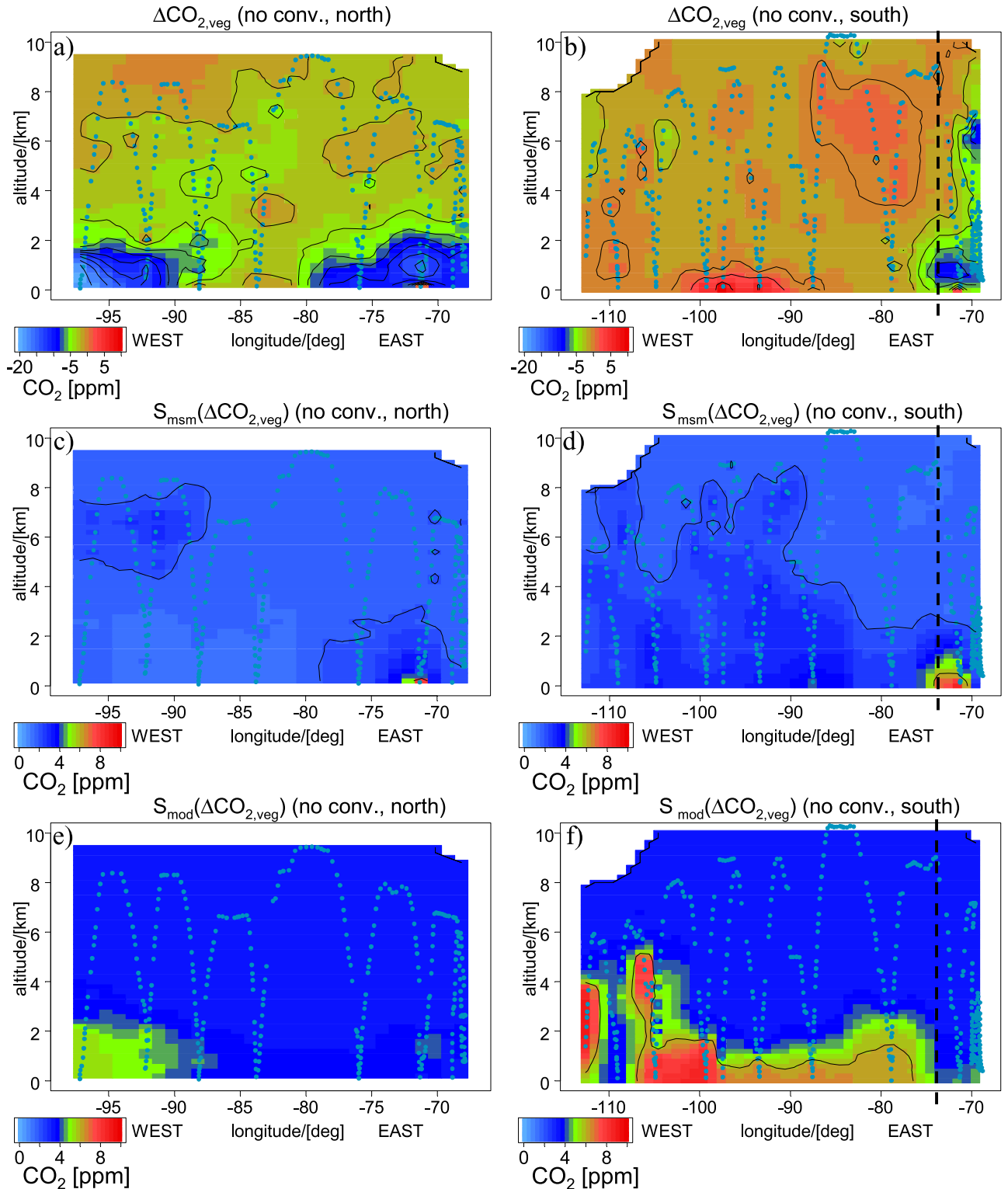
Figure 7. Measurement derived CO_2 vegetation signals after equation (9), for the (a) northern survey and (b) southern survey, (c and d) uncertainty of the CO_2 vegetation signal, and (e and f) uncertainty of the modeled vegetation signal due to the model error components S_{part}, S_{eddy}, S_{transp}, S_{aggr}, and S_{ocean}. All results are shown for the non-convective case. The vertical dashed line in Figure 7b marks the eastern end of the data used in the optimization for the southern survey.
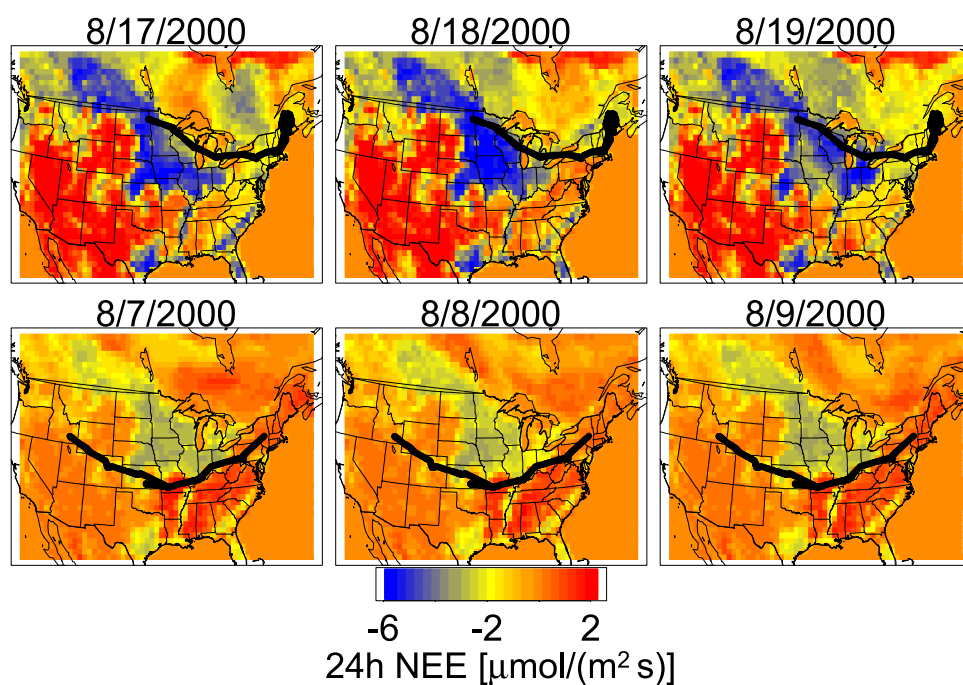
Figure 9. Twenty-four-hour averages of NEE from forests and croplands based on the upscaling of AmeriFlux data for three consecutive days preceding the measurements (results for non-convection case). The top and bottom rows show results using scaling factors constrained by the northern and the southern surveys, respectively.
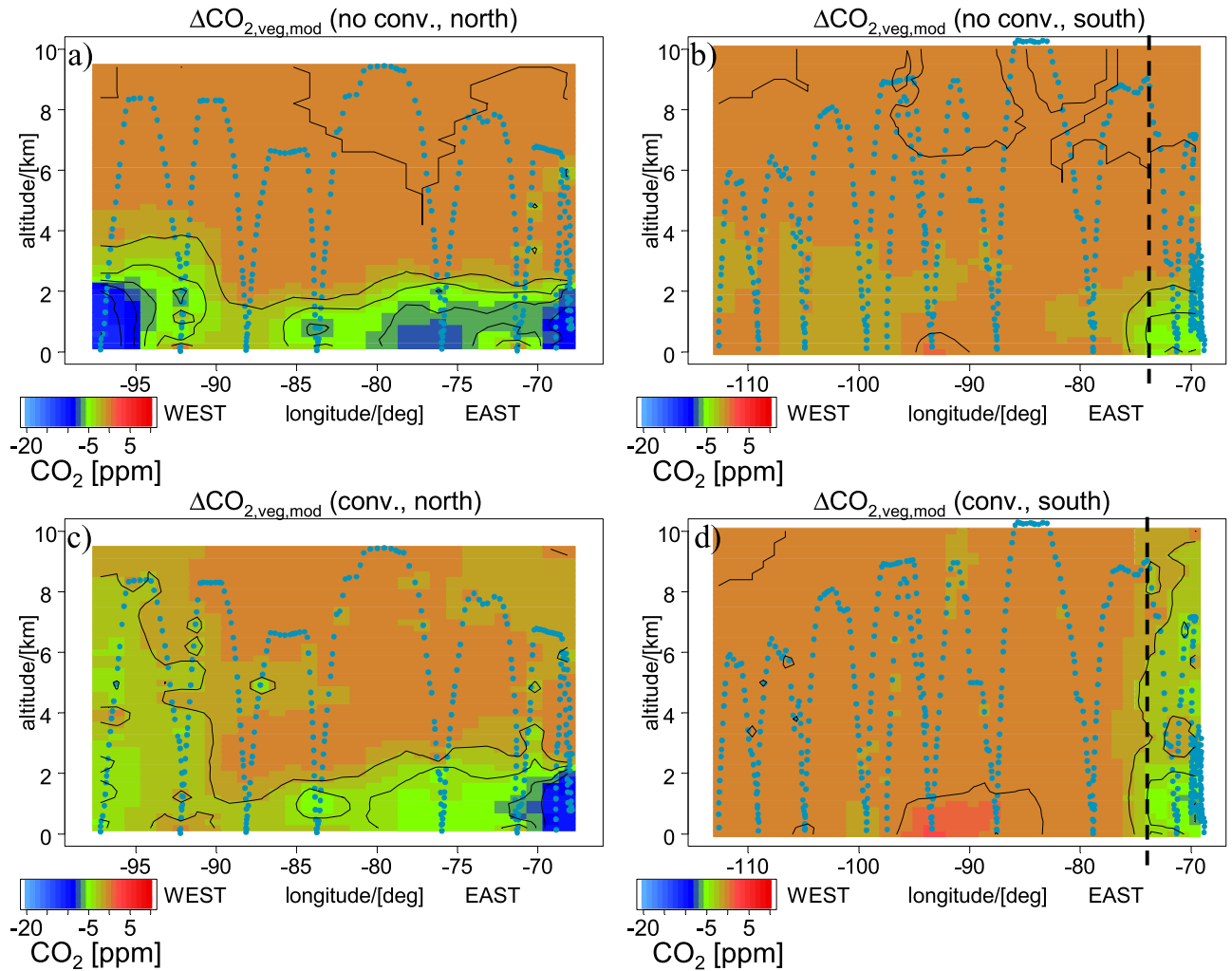
Figure 11. Modeled vegetation CO2 signal (a and b) for the non-convective case and (c and d) for the convective case. Left column (Figures 11a and 11c) shows results for the northern survey, right column (Figures 11b and 11d) for the southern survey. Left of the vertical dashed lines in Figures 11b and 11d the southern survey scaling factors are used; to the right of these lines, northern survey scaling factors were applied.