SI Appendix

S1 Flux Module Descriptions

Annual global total fossil fuel CO_2 emissions are from Marland *et al.* [1] which extend through 2003 and are then linearly extrapolated through 2005. Fluxes are spatially distributed according to the EDGAR inventories [2] and include a seasonal cycle based on the Blasing *et al.* [3] analysis for the United States. Carbon emissions in the form of CO, VOC's, and CH_4 are not accounted for separately in this work possibly leading to slightly higher uptake [4]. Monthly fire CO_2 emissions are taken from the Global Fire Emissions Database version 2 (GFED2) [5, 6]. Ocean CO_2 exchange is based on $5^\circ \times 4^\circ$ ΔpCO_2 estimates from Takahashi *et al.* [7], combined with 3-hourly surface pressure and 10m wind speed from the ECMWF model to give synoptic variations as well as inter-annual variability identical to the procedures described in Kettle *et al.* [8]. Finally, NEP is created from the monthly mean CASA Net Primary Production (NPP) and heterotrophic respiration (R_H) that include the effect of fire on the CASA simulated carbon pools in vegetation and soils, as well as recovery of the ecosystem after burning. Higher frequency variations (diurnal, synoptic) are added to these fluxes every 3 hours using a simple temperature Q_{10} relationship and linear scaling with solar radiation, similar to the procedure in Olsen and Randerson [9].

The Olson ecosystem classification used to scale the fine scale patterns of the fluxes above is summarized in Table 1. The scaling factors across ecoregions currently have limited physical realism for a quite simple reason: we do not know which physical parameter in the terrestrial biosphere controls NEP across larger areas and can be constrained from the atmosphere. Finding this parameter is a problem for all carbon cycle scientists attempting to scale NEP information beyond their local scope and multiple studies suggest that there is likely not one such parameter controlling NEP, but a large set of them working on different sides of the photosynthesis and respiration balance. The linear scaling factors we chose absorb anything that might be wrong with the NEP magnitude from the underlying terrestrial carbon cycle model: lack of human management component, poorly initialized carbon pools, dependence on noisy NDVI, and scaling of 3-hourly fluxes with a simple Q10.

The choice to take linear scaling factors across a whole ecosystem furthermore assumes that the relative NEP patterns one $1^\circ \times 1^\circ$ produced by the terrestrial biosphere model are correct within that ecosystem. These patterns currently depend mostly on the forcing of the CASA model of which high resolution meteorology is one. Through this, regional and interannual variations in climate patterns across an ecosystem such as the boreal forests are represented in our fluxes through CASA's grid-box specific response, but the multiplication with the derived scaling factor will not alter the predicted pattern of fluxes, only its magnitude in NEP across the whole ecosystem. Inherent in this a-priori combination of regional quantities such as fluxes into larger areas is the aggregation error described by Kaminski *et al.*[10]. Sensitivity test M2 addresses some of this potential error in anticipation of higher resolution studies in the near future. Note that as the observing network expands, we expect to be able to optimize for smaller ecoregions and thereby relax the assumption on correct NEP patterns across large regions.

S1.1 2005 Input Data

For the last year of our inversion (2005) some data sets were not available in time to include in our final results. This includes for instance the GFED2 weekly fire product through 2005 available now,

Table 1: Ecosystem types considered for terrestrial fluxes. The land-surface characterization is based on Olson et al, 1985 [11] and each $1^\circ \times 1^\circ$ gridbox is assigned to one category based on the locally dominant vegetation type. Percentages indicate the area associated with each category for North America

category #	Description
1	Conifer Forest (19.0%)
2	Broadleaf Forest (1.3%)
3	Mixed Forest (7.5%)
4	Grass/Shrub (12.6%)
5	Tropical Forest (0.3%)
6	Scrub/Woods (2.1%)
7	Semitundra (19.4%)
8	Fields/Woods/Savanna (4.9%)
9	Northern Taiga (8.1%)
10	Forest/Field (6.3%)
11	Wetland (1.7%)
12	Deserts (0.1%)
13	Shrub/Tree/Suc (0.1%)
14	Crops (9.7%)
15	Conifer Snowy/Coastal (0.4%)
16	Wooded tundra (1.7%)
17	Mangrove (0.0%)
18	Ice and Polar desert (0.0%)
19	Water (4.9%)

instead of the monthly one through 2004 that we used. Our 2005 fire emissions are the climatological average of those used in 2000-2004. Also, the CO_2 measurements obtained from the Environment Canada (EC) (sites FRD_06C0, ALT_06C0, OBS_06C0) currently only go through February 2005. The 2005 fossil fuel emissions are identical to those used in 2004. Although economic statistics are now available to extrapolate the fossil fuel emissions from previous years, we did not include this information at the time we created the results. The month December of 2005 was run without the United States $1^\circ \times 1^\circ$ zoom region in the transport model. This is due to a change in model resolution in the parent model from ECMWF in February of 2006. As a result of these caveats, 2005 currently has higher variances than 2004 and our estimate will likely be refined in the near future.

Full documentation of the flux modules used in consecutive CarbonTracker releases can be found at http://carbontracker.noaa.gov

S2 Observations

At the continuous sampling sites, we construct one daytime average (12:00-16:00 Local Time) mixing ratio for each day in the time series, recognizing that our atmospheric transport model does not always captures the continental nighttime stability regime while daytime well-mixed conditions are better matched. This approach is partly based on analysis of TransCom Continuous results (Law et al, *manuscript submitted to GBC*, 2007). Moreover, observations at sub-daily time scales are likely to be strongly correlated and therefore add relatively little independent information to our results. Also based on TransCom Continuous simulations, we decided to move a set of coastal sites by one degree into the ocean to force the model sample to be more representative of the actual site conditions. These sites are labeled for reference in the complete table of sites used in our study (Table 2). Note that experiments Obs2 and Obs3 in Section S4 address the fairly small sensitivity of our results to the choice of network.

We apply a further selection criterion during the assimilation to exclude non-Marine Boundary Layer (MBL) observations that are very poorly forecast in our framework. We interpret an observed-minus-forecasted (OmF) mixing ratio that exceeds 3 times the prescribed model-data mismatch as an indicator that our modeling framework fails. This can happen for instance when an air sample is representative of local exchange not captured well by our $1^\circ \times 1^\circ$ fluxes, when local meteorological conditions are not captured by our offline transport fields, but also when large-scale CO_2 exchange is suddenly changed (e.g. fires, pests, droughts) to an extent that can not be accommodated by our flux modules. This last situation would imply an important change in the carbon cycle and has to be recognized when analyzing the results. In accordance with the 3-σ rejection criterion, $\sim$2% of the observations are discarded through this mechanism in our assimilations. Table 2 gives a summary of the sites used and the assimilation performance.

To assign model-data mismatches, we divided all observation sites into six categories. These categories and respective model-data mismatches [ppm] are: marine boundary layer (0.75), continental sites (2.5), mixed land/ocean and mountain sites (1.50), continuous sites (3.0), and difficult sites (7.5). These values represent subjective choices and are not based on an optimization or analysis of representation errors in our model. However, sites were categorized to yield an innovation χ^2 close to 1.0 in each category, for which the six absolute magnitudes were manually tuned to achieve this as well. We tested one alternative for model-data mismatch in Section S4 (experiment R2) which yielded higher posterior covariances and a poorer match to observations, as would be expected. Figure 1 shows the location of all sites in Table 1. Figure 2 shows three examples of

observed CO_2 time series and their simulated counterparts resulting from the assimilation. A full set of these figures for all sites listed in Table 1 is available from http://carbontracker.noaa.gov. Figure 3 shows a similar comparison for independent CO_2 mole fraction profiles measured from a small aircraft over Poker Flat, Alaska, USA.

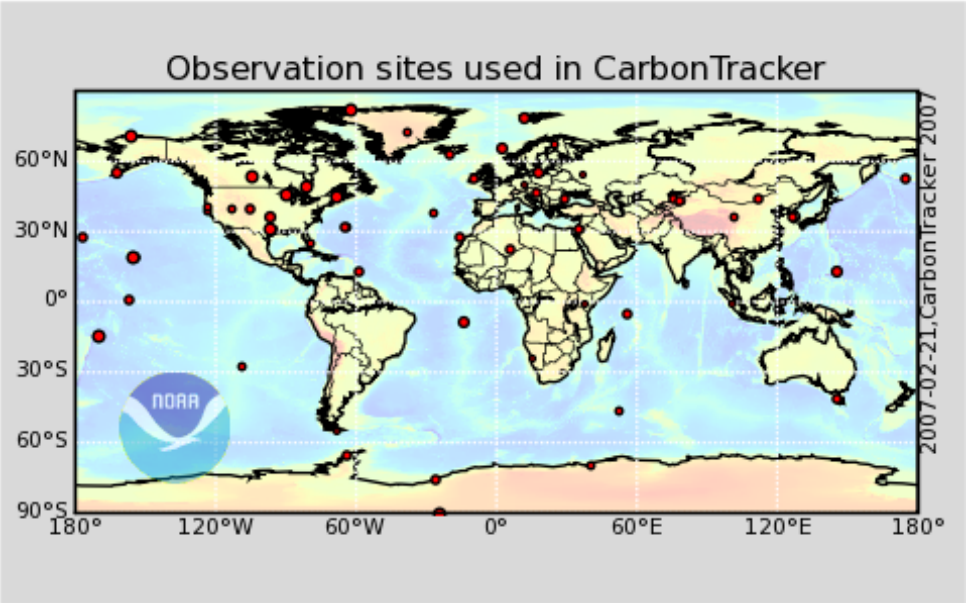

Figure 1: Locations of sites listed in Table 1.

Table 2: Sites and samples used in this study, their assigned model-data mismatch category, and their innovation χ^2 statistic. Sites marked with an asterisk have been moved by 1° longitude and/or latitude in the model to sample more representative air masses. The elevation of a site denotes its height in meters above sea level and does not include the vertical extension of the intake height.

Code	Name	lat,lon,elev	Lab	$N^{(1)}$	$M^{(2)}$	$r^{(3)}$	$\chi^{2(4)}$
		Continuous Analyzers					
alt_06C0	Alert, Nunavut, Canada	82 27'N, 62 31'W, 200.0m	MSC	1737	0	2.50	0.20
amt_01C3	Argyle, Maine, United States	45 2'N, 68 41'W, 50.0m	NOAA	654	52	3.00	0.57
brw_01C0*	Barrow, Alaska, United States	71 19'N, 156 36'W, 11.0m	NOAA	1765	2	2.50	0.27
frd_06C0	Fraserdale, Canada	49 53'N, 81 34'W, 210.0m	MSC	1711	40	3.00	0.42
mlo_01C0	Mauna Loa, Hawaii, United States	19 32'N, 155 35'W, 3397.0m	NOAA	1197	0	0.75	0.89
obs_06C0	Candle Lake, Saskatchewan, Canada	53 59'N, 105 7'W, 629.0m	MSC	687	4	3.00	0.55
smo_01C0	Tutuila, American Samoa	14 14'S, 170 34'W, 42.0m	NOAA	1965	0	0.75	0.66
spo_01C0	South Pole, Antarctica, United States	89 59'S, 24 48'W, 2810.0m	NOAA	2074	0	0.75	0.45
wkt_01C3	Moody, Texas, United States	31 19'N, 97 20'W, 251.0m	NOAA	544	26	3.00	0.54
lef_01C3	Park Falls, Wisconsin, United States	45 56'N, 90 16'W, 472.0m	NOAA	1532	31	3.00	0.54
		Flask Samples					
alt_01D0	Alert, Nunavut, Canada	82 27'N, 62 31'W, 200.0m	NOAA	279	0	1.50	0.45
asc_01D0	Ascension Island, United Kingdom	7 55'S, 14 25'W, 54.0m	NOAA	499	0	0.75	1.31
ask_01D0	Assekrem, Algeria	23 11'N, 5 25'E, 2728.0m	NOAA	258	0	1.50	0.39
azr_01D0	Terceira Island, Azores, Portugal	38 46'N, 27 23'W, 40.0m	NOAA	189	4	1.50	0.91
bal_01D0	Baltic Sea, Poland	55 21'N, 17 13'E, 3.0m	NOAA	447	1	7.50	0.45
bkt_01D0	Bukit Kototabang, Indonesia	0 12'S, 100 19'E, 864.5m	NOAA	74	0	7.50	0.55
bme_01D0	St. Davids Head, Bermuda, United Kingdom	32 22'N, 64 39'W, 30.0m	NOAA	189	9	1.50	1.26
bmw_01D0	Tudor Hill, Bermuda, United Kingdom	32 16'N, 64 53'W, 30.0m	NOAA	203	3	1.50	1.00
brw_01D0	Barrow, Alaska, United States	71 19'N, 156 36'W, 11.0m	NOAA	269	6	1.50	0.65
bsc_01D0*	Black Sea, Constanta, Romania	44 10'N, 28 41'E, 3.0m	NOAA	247	1	7.50	0.75
cba_01D0*	Cold Bay, Alaska, United States	55 12'N, 162 43'W, 25.0m	NOAA	438	22	1.50	1.18
cgo_01D0*	Cape Grim, Tasmania, Australia	40 41'S, 144 41'E, 94.0m	NOAA	245	0	1.50	0.20
chr_01D0	Christmas Island, Republic of Kiribati	1 42'N, 157 10'W, 3.0m	NOAA	223	0	0.75	1.19
crz_01D0	Crozet Island, France	46 27'S, 51 51'E, 120.0m	NOAA	180	0	0.75	0.96
eic_01D0	Easter Island, Chile	27 9'S, 109 27'W, 50.0m	NOAA	129	0	7.50	0.04
gmi_01D0	Mariana Islands, Guam	13 26'N, 144 47'E, 1.0m	NOAA	463	0	1.50	0.47
hba_01D0	Halley Station, Antarctica, United Kingdom	75 35'S, 26 30'W, 30.0m	NOAA	274	0	0.75	0.89

Continued on Next Page...

Code	Name	lat,lon,elev	Lab	$N^{(1)}$	$M^{(2)}$	$r^{(3)}$	$\chi^{2(4)}$
hun_01D0	Hegyhatsal, Hungary	46 57'N, 16 39'E, 248.0m	NOAA	274	3	7.50	0.44
ice_01D0*	Storhofdi, Vestmannaeyjar, Iceland	63 20'N, 20 17'W, 118.0m	NOAA	258	2	1.50	0.52
izo_01D0	Tenerife, Canary Islands, Spain	28 18'N, 16 29'W, 2360.0m	NOAA	209	2	1.50	0.92
key_01D0*	Key Biscayne, Florida, United States	25 40'N, 80 12'W, 3.0m	NOAA	196	0	2.50	0.31
kum_01D0	Cape Kumukahi, Hawaii, United States	19 31'N, 154 49'W, 3.0m	NOAA	266	0	1.50	0.49
kzd_01D0	Sary Taukum, Kazakhstan	44 27'N, 75 34'E, 412.0m	NOAA	271	55	2.50	0.65
kzm_01D0	Plateau Assy, Kazakhstan	43 15'N, 77 53'E, 2519.0m	NOAA	238	1	2.50	1.13
mhd_01D0*	Mace Head, County Galway, Ireland	53 20'N, 9 54'W, 25.0m	NOAA	224	0	2.50	0.23
mid_01D0	Sand Island, Midway, United States	28 13'N, 177 23'W, 3.7m	NOAA	265	0	1.50	0.62
mkn_01D0	Mt. Kenya, Kenya	0 3'S, 37 18'E, 3897.0m	NOAA	57	0	2.50	1.00
mlo_01D0	Mauna Loa, Hawaii, United States	19 32'N, 155 35'W, 3397.0m	NOAA	310	0	1.50	0.28
nmb_01D0	Gobabeb, Namibia	23 35'S, 15 2'E, 456.0m	NOAA	16	0	2.50	0.17
nwr_01D0	Niwot Ridge, Colorado, United States	40 3'N, 105 35'W, 3523.0m	NOAA	264	2	1.50	0.74
obn_01D0	Obninsk, Russia	55 7'N, 36 36'E, 183.0m	NOAA	72	0	7.50	0.42
oxk_01D0	Ochsenkopf, Germany	50 4'N, 11 48'E, 1193.0m	NOAA	19	1	2.50	1.07
pal_01D0	Pallas-Sammaltunturi, GAW Station, Finland	67 58'N, 24 7'E, 560.0m	NOAA	145	1	2.50	0.82
poc_01D1	Pacific Ocean, N/A	99 59'S, 999 59'W, 10.0m	NOAA	896	1	7.50	0.01
psa_01D0	Palmer Station, Antarctica, United States	64 55'S, 64 0'W, 10.0m	NOAA	276	0	0.75	1.16
pta_01D0*	Point Arena, California, United States	38 57'N, 123 44'W, 17.0m	NOAA	180	0	7.50	0.34
rpb_01D0	Ragged Point, Barbados	13 10'N, 59 26'W, 45.0m	NOAA	267	0	1.50	0.72
sey_01D0	Mahe Island, Seychelles	4 40'S, 55 10'E, 3.0m	NOAA	256	0	0.75	1.26
sgp_01D0	Southern Great Plains, Oklahoma, United States	36 48'N, 97 30'W, 314.0m	NOAA	369	29	2.50	0.49
shm_01D0	Shemya Island, Alaska, United States	52 43'N, 174 6'E, 40.0m	NOAA	234	1	2.50	0.82
smo_01D0	Tutuila, American Samoa	14 14'S, 170 34'W, 42.0m	NOAA	315	0	1.50	0.18
spo_01D0	South Pole, Antarctica, United States	89 59'S, 24 48'W, 2810.0m	NOAA	291	0	1.50	0.09
stm_01D0	Ocean Station M, Norway	66 0'N, 2 0'E, 0.0m	NOAA	504	5	1.50	0.76
sum_01D0	Summit, Greenland	72 35'N, 38 29'W, 3238.0m	NOAA	178	0	1.50	0.46
syo_01D0	Syowa Station, Antarctica, Japan	69 0'S, 39 35'E, 11.0m	NOAA	139	0	0.75	1.34
tap_01D0*	Tae-ahn Peninsula, Republic of Korea	36 44'N, 126 8'E, 20.0m	NOAA	202	0	7.50	0.38
tdf_01D0	Tierra Del Fuego, Ushuaia, Argentina	54 52'S, 68 29'W, 20.0m	NOAA	74	0	0.75	0.67
thd_01D0*	Trinidad Head, California, United States	41 3'N, 124 9'W, 107.0m	NOAA	134	0	7.50	0.44
uta_01D0*	Wendover, Utah, United States	39 54'N, 113 43'W, 1320.0m	NOAA	250	55	2.50	0.32
uum_01D0	Ulaan Uul, Mongolia	44 27'N, 111 6'E, 914.0m	NOAA	273	5	2.50	0.67
wis_01D0	Sede Boker, Negev Desert, Israel	31 8'N, 34 53'E, 400.0m	NOAA	296	1	2.50	0.79
wlg_01D0	Mt. Waliguan, Peoples Republic of China	36 17'N, 100 54'E, 3810.0m	NOAA	166	6	1.50	1.16
zep_01D0	Ny-Alesund, Svalbard, Norway and Sweden	78 54'N, 11 53'E, 475.0m	NOAA	355	2	1.50	0.70

Continued on Next Page...

Code	Name	lat,lon,elev	Lab	N[1]	M[2]	r[3]	$\chi^{2(4)}$
	Total			27675	373	0.00	0.55

[1]Denotes total number of observations available
[2]Denotes total number of observations flagged (see text)
[3]The assigned model-data mismatch in [ppm]
[4]Calculated from the innovations as in Peters et al. [2005]

S3 The Data Assimilation System

The ensemble system used to solve for the scalar multiplication factors is similar to that in Peters *et al.* [12]. In this work, we have restricted the length of the smoother window to only five weeks as we found the derived flux patterns within North America to be robustly resolved well within that time (experiments L2, L3 in Section S4). Longer window lengths tended to increase covariations between parameters without substantial further decrease of the estimated variances or changes to the parameter values. Moreover, tests with our transport model indicated that regional flux signals had diffused over much of the Northern Hemisphere after five weeks, thus leaving almost no useful gradients to infer sub-continental patterns. Finally, the small remaining signal after five weeks was found to be relatively sensitive to the formulation of vertical exchange (diffusion and convection), suggesting a possible large transport error component after this time, and the danger of biasing our results. We caution the reader that although the North American flux results were found to be robust after five weeks, regions of the world with less dense observational coverage (tropics, Southern Hemisphere, parts of Asia) are likely to be poorly observable even after more than a month of transport and therefore less robustly resolved. Although longer assimilation windows, or long prior covariance length-scales, could potentially help to 'observe' larger scale emission totals from such areas, we focus our analysis here on a region more directly constrained by atmospheric observations.

S3.1 Lag, members, localization

Ensemble statistics are created from 150 ensemble members, each with their own background CO_2 concentration field to represent the time history (and thus covariances) of the filter. Experiment N2 in Section S4 used twice as many ensemble members with no significant effect on the fluxes. To dampen spurious noise due to the approximation of the covariance matrix, we apply localization [13] for non-MBL sites only. This ensures that tall-tower observations within North America do not inform on for instance tropical African fluxes, unless a very robust signal is found. In contrast, MBL sites with a known large footprint and strong capacity to see integrated flux signals are not localized.

S3.2 Background parameter values and dynamical model

In this work similar to Peters *et al* [12], the dynamical model is applied to the mean parameter values λ as:

$$\lambda^{\mathrm{b}}_{\mathrm{t}} = (\lambda^{\mathrm{a}}_{\mathrm{t-2}} + \lambda^{\mathrm{a}}_{\mathrm{t-1}} + \lambda^{\mathrm{p}})/3.0 \tag{1}$$

Where superscript a refers to analyzed quantities from previous steps, superscript b refers to the background values for the new step, and superscript p refers to real a-priori determined values that are fixed in time and chosen as part of the inversion set-up. Physically, this model describes that parameter values λ for a new time step are chosen as a combination between optimized values from the two previous time steps, and a fixed prior value. This operation is similar to the simple persistence forecast used previously [12], but represents a smoothing over three time steps thus dampening variations in the forecast of λ^b in time. The inclusion of the prior term λ^p acts as a regularization [14] and ensures that the parameters in our system will eventually revert back to predetermined prior values when there is no information coming from the observations. Note that our dynamical model equation does not include an error term on the dynamical model, for the simple

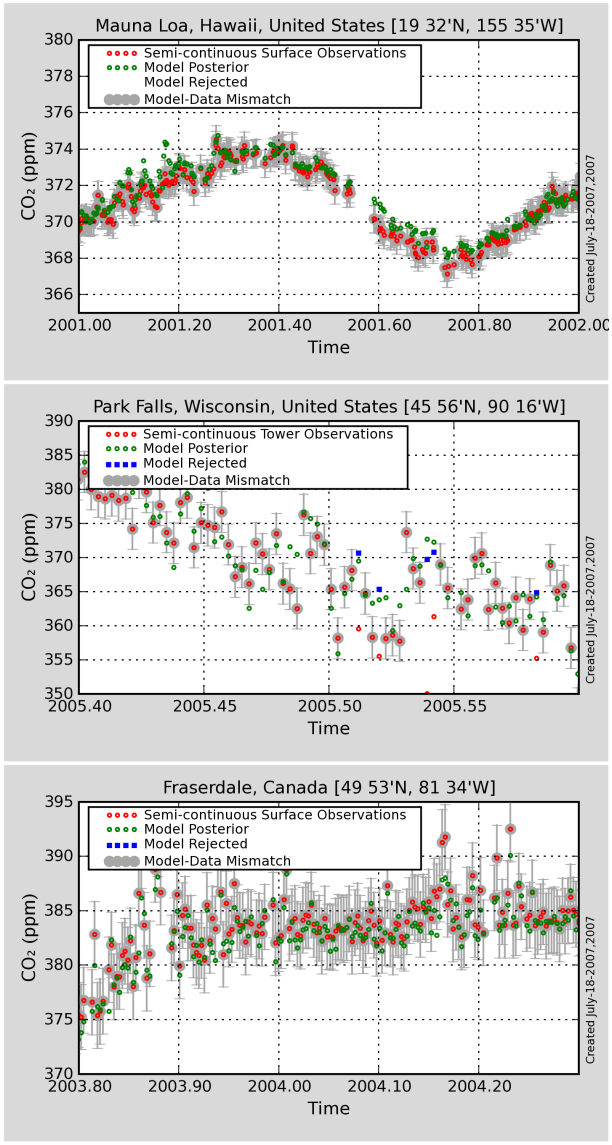

Figure 2: Observed and modeled time series of CO_2 mole fractions for a small subset of sites used in the assimilation. Modeled CO_2 is co-sampled from the optimized fluxes run forward through TM5. Note that each figure covers a different time period to show extra detail. A full set of figures for all sites listed in Table 1 can be found at http://carbontracker.noaa.gov. Fraserdale data is courtesy of Environment Canada.

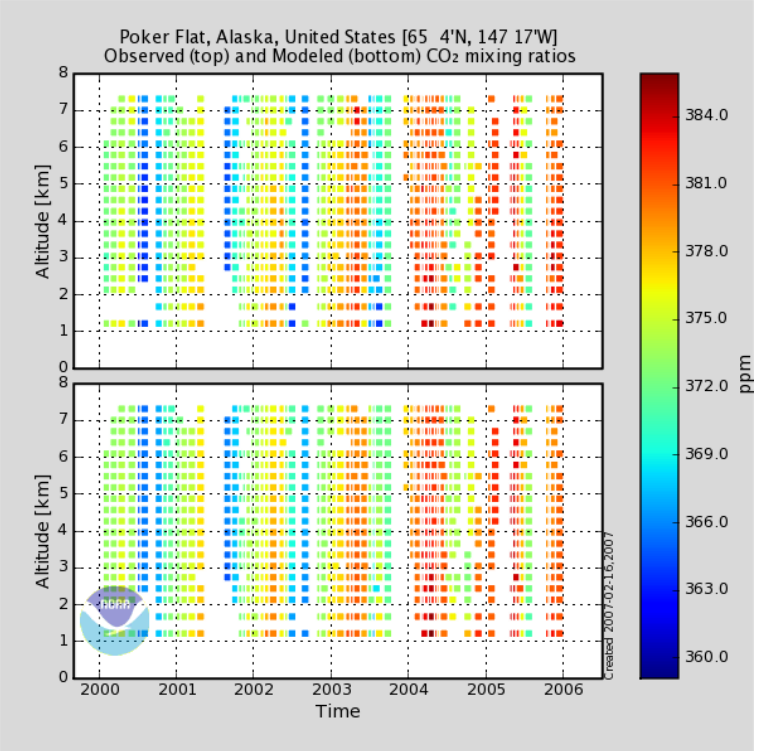

Figure 3: Observed and modeled time series of CO_2 mole fractions for Poker Flat, Alaska aircraft vertical profiles. Aircraft sites were not used in the assimilation. Modeled CO_2 is co-sampled from the optimized fluxes run forward through TM5. A full set of figures for all aircraft sites can be found at http://carbontracker.noaa.gov.

reason that we don't know the error of this model. This is reflected in the treatment of covariance, which is always set to a prior covariance structure and not forecasted with our dynamical model.

Our choice to use a combination of 3 states is a compromise between prescribing prior fluxes at each step and letting the system propagate all information from one step to the next without any guidance. Although the latter will work fine for the North American fluxes which are strongly constrained by observations, fluxes in most of the rest of the world need the regularization to stay within reasonable bounds in the absence of observational information and a proper dynamical model for CO_2 fluxes. The posterior covariance can be used to diagnose areas where minimal observational constraints necessitate the use of the prior term. We furthermore note that the described dynamical model imposes a smoothness on the estimated state vectors similar to a prior temporal covariance structure, which we do not specify in our system. Next, we will describe the chosen values of λ^p.

S3.3 Covariances

Prior values for λ^p are all 1.0 to yield fluxes that are unchanged from their values predicted in our modules. The prior covariance structure $\mathbf{P}^p$ describes the magnitude of the uncertainty on each parameter, plus their correlation in space. The latter is applied such that the same ecosystem types in different TransCom regions decrease exponentially with distance (L=2000km), and thus assumes a coupling *between* the behavior of the same ecosystems in close proximity to one another (such as coniferous forests in Boreal and Temperate North America). Furthermore, all ecosystems *within* tropical TransCom regions are coupled decreasing exponentially with distance since we do not believe the current observing network can constrain tropical fluxes on sub-continental scales, and want to prevent large dipoles to occur in the tropics.

In our standard assimilation, the chosen standard deviation is 80% on land parameters, and 40% on ocean parameters. This reflects more prior confidence in the ocean fluxes than in terrestrial fluxes, as is assumed often in inversion studies and partly reflects the lower variability and larger homogeneity of the ocean fluxes. All parameters have the same variance within the land or ocean domain. Because the parameters multiply the net-flux though, ecosystems with larger weekly mean net fluxes have a larger variance in absolute flux magnitude. Experiments C2, C3 test alternatives to the prescribed covariance structure and show some sensitivity especially to the choice of ocean uncertainty.

S4 Uncertainty Estimates

Although our framework returns formal covariance estimates for each week of optimization, our approach to prescribe the background covariances at each step and draw from them randomly means that temporal covariations over time scales of more than our assimilation window are not retained. Annual mean covariance estimates therefore do not reflect the real uncertainty of our annual mean fluxes and would not do justice to the tight match to over 27,000 observations we have achieved. Even if the temporal covariations were available though, they would only reflect the random component of the particular minimization problem we constructed, rather than a characterization of the true uncertainties of the full assimilation system. It has been shown previously that, if sufficient volumes of data is available, the random component becomes much smaller than the systematic error component [15, 16].

As an alternative, we have conducted a set of sensitivity experiments aimed at characterizing the spread in the flux estimates when taking different yet plausible approaches to the design of

the assimilation system. Although this set is not comprehensive due to computational limitations, it reflects the influence of certain fundamental choices one faces when designing a system like CarbonTracker. We use the minimum and maximum flux values found in this set to denote the range in which we expect the true flux to be. Our best estimate presented in the main text represents the scenario we found most suitable based on an analysis of the input data itself, as well as the comparison to assimilated and independent observations.

The alternative assimilation experiments for the period 2000-2002 used a coarse $6^\circ \times 4^\circ$ degree version of the TM5 transport model. Due to the cost of running our assimilation system (9 CPU months for 2 years of assimilation), longer tests and more axis of variability were not assessed at this point. The base run for this set of tests was a coarse TM5 run with ocean and fossil fuel fluxes as in the main text, a neutral biosphere run from CASA as biological priors, and model-data mismatches slightly higher than used in the final run. All sensitivity runs varied one aspect of the assimilation setup.

Table 3 shows the estimated fluxes for the North American ecosystem from each simulation. Note that all sensitivity runs except T2 were done at $6^\circ \times 4^\circ$, and we present results from 2001 only. Also note that the transport model itself is one of the largest sensitivities in flux inversions, which was not varied in this study and thus not included in our range of estimates. Future work will try to address this component in more detail.

B4: Biosphere flux modules follow monthly mean NEE from a GFED2 biosphere run of the CASA model, combined with 3-hourly solar radiation and temperature from the TM5 model, GFED2 fires are included
B2: Biosphere flux modules follow monthly mean NEE from a neutral biosphere run of the CASA model, no fires are included
M2: Parameters λ follow a terrestrial ecosystem specification with 19 categories defined following the ECMWF land-surface characterization. The ECMWF classification is based on the Global Land Cover Characteristics (GLCC) data (http://edcdaac.usgs.gov/glcc/glcc.html) [17]. For comparison purposes, ECMWF categories were approximately mapped to Olson categories.
O2: Ocean fluxes are 3-hourly CO_2 based on monthly mean CO_2 fluxes from the [18] joint ocean-atmosphere inversion combined with ECMWF wind speed and surface pressure variability
C2: The standard deviation in the covariance matrix is 160% on land parameters, and 80% on ocean parameters
C3: The standard deviation in the covariance matrix is 80% on land parameters, and 80% on ocean parameters
R2: Model-data mismatch values are doubled
L2: Ensemble system runs with 3 weeks of lag
L3: Ensemble system runs with 10 weeks of lag
N2: Ensemble system runs with 300 ensemble members
Obs2: 50% of available observations are discarded at random each week
Obs3: Daily average mixing ratios from observatories and tall-towers around North America are not included
F2: Fossil Fuel emission patterns based on Andres *et al* [20], scaled to 2000-2005 global total emissions from EDGAR fast track without any seasonal variations
T2: The TM5 transport model runs at global $6^\circ \times 4^\circ$, and two nested grids

Note that the best estimate, using a combination of settings from all experiments does not

necessarily fall in the middle of the estimated range but might be near the edge of the distribution. Because the estimates of the mean furthermore span the full six year period whereas the uncertainty is for 2001 only, the best estimate might even be outside the range for some of the smaller components.

Based on these results, lower model-data mismatches, higher transport model resolution, and GFED2 prior fluxes for the biosphere and fires were included in the final results. The sensitivity tests show that the results presented here are sensitive, but not in a systematic direction, against a large number of changes, and that the setup chosen for our 6-year inversion is unlikely to have biased annual mean results relative to other credible choices. Largest sensitivity in flux results is to the choice of high resolution prior flux patterns (O2,B2,B4) and the TM5 model resolution (T2). The runs with higher resolution (T2) and fires (B4) included provide a notably better match to the full set of observations, whereas the different ocean (O2) prior and the use of monthly mean fluxes (B2) only improved results at selected sites. North American flux results were largely insensitive to the configuration of the inversion (L3, N2, R2, C2, C3) as long as the lag of the filter was not too short (L2). The consistency of the results even when half of the observations were discarded (Obs2) shows adequate redundancy in the information from different sites. The exclusion of continuous data (Obs3) did not change our estimate significantly, suggesting that the two sets of observations can safely be mixed without biasing the results towards one set. Note however that for the sensitivity experiments the tower and flask sites were balanced by their model-data mismatch to contribute equally, whereas tower sites were given higher relative weight in the final runs. Thus, it should not be concluded that tower data do not add information over the regular flask network from these tests.

Table 3: Results of the sensitivity experiments conducted in this study (see text). One aspect of the model was varied per sensitivity run. The table includes only the Olson categories with significantly large fluxes (>0.05 PgC/yr).

Olson Category	1	3	4	7	8	9	10	14	Total
B4	-0.38	-0.11	-0.09	-0.09	-0.05	-0.09	-0.08	-0.08	-1.01
B2	-0.37	-0.07	-0.05	-0.07	-0.02	-0.07	-0.10	-0.09	-0.84
L2	-0.22	-0.05	-0.03	-0.09	-0.01	-0.06	-0.07	-0.12	-0.66
R2	-0.19	-0.05	-0.03	-0.06	-0.03	-0.03	-0.1	-0.11	-0.60
C3	-0.19	-0.03	-0.04	-0.08	-0.04	-0.03	-0.08	-0.07	-0.58
O2	-0.19	-0.09	-0.02	-0.07	-0.02	-0.05	-0.09	-0.06	-0.58
C2	-0.13	-0.05	-0.01	-0.10	-0.01	-0.07	-0.10	-0.10	-0.57
M2[a]	-0.20[b]	-0.05[c]	-0.04[d]	-0.07[e]	-0.04[f]	-0.04[g]	-0.03[h]	-0.12[i]	-0.56
Obs2	-0.27	-0.02	-0.04	-0.06	0.01	-0.04	-0.09	-0.04	-0.56
L3	-0.14	-0.02	-0.02	-0.07	-0.03	-0.05	-0.10	-0.10	-0.53
N2	-0.18	-0.06	-0.00	-0.09	-0.00	-0.04	-0.07	-0.08	-0.52
Obs3	-0.18	-0.01	-0.03	-0.06	-0.02	-0.03	-0.07	-0.12	-0.52
F2	-0.16	-0.04	-0.02	-0.08	-0.04	-0.03	-0.06	-0.04	-0.49
T2	-0.05	0.00	-0.06	-0.10	-0.05	-0.05	-0.05	-0.03	-0.40

[a]Olson *et al* categories were approximately matched by similar ecosystems in the Loveland *et al.* database
[b]Coniferous Forest = Evergreen Needleleaf forest
[c]Mixed Forest = Mixed Forest
[d]Grass/Shrub = Short Grass + Tall Grass + Evergreen Shrub
[e]Semi Tundra = Tundra
[f]Fields/Woods/Savanna = Interrupted Forest (but large fraction overlaps with ECMWF Crops)
[g]Northern Taiga = Deciduous shrubs + Bogs&Marshes
[h]Forest/Field = Evergreen Broadleaf (but large fraction overlaps with ECMWF Crops)
[i]Crops = Crops, Mixed farming (but fraction of ECMWF Crops includes other Olson categories)

S5 Global Flux Results

Fluxes for the rest of the globe are presented in Table 4, and aggregated to annual means for TransCom regions. For reference, we have included the annual means from Gurney *et al* [21], Baker *et al* [14], and Jacobson *et al* [18] for the same areas. Note that these estimates span a different time period and were estimated with a different inversion technique. Nevertheless, results generally correspond quite well because there is similarity in the observations used, the flux prior models, and even the transport models. Largest differences can be found in the tropical regions, and in the partitioning of the NH land sink between Europe, North America, and the Asia. The small tropical fluxes coupled to a smaller NH land sink are unique to our inversion and part of ongoing more detailed analysis.

The higher resolution flux estimate by Rödenbeck *et al.* [19] spans the period 1982-2001 based on some of the same flask data used in this study. Although the long-term mean North American

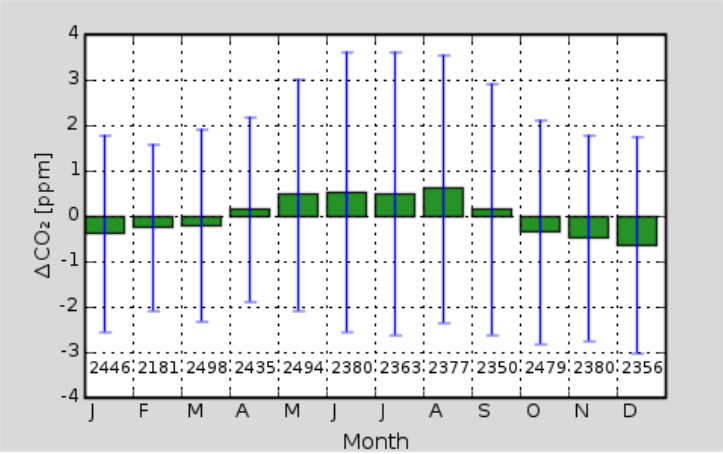

Figure 4: Modeled minus observed CO_2 for a set of 28,000 observations assimilated. The differences are averaged by month and the bar indicates the standard deviation on the monthly mean. The number below each bar denotes the number of observations used in the statistic. Units are ppm.

terrestrial uptake found by these authors is similar to our estimated uptake (∼0.5 PgC/yr), the flux in the overlapping years 2000-2001 does not agree between the two studies. A sudden change to positive carbon fluxes in Rödenbeck *et al.* [19] during this period is ascribed to the data from one single site, Key Biscayne (see their Table 7). This feature disappears when Key Biscayne is not included and the authors do not see this results as real behavior of the carbon cycle. It demonstrates the sensitivity of their setup to individual sites, brought about by the relatively small number of observations ingested. CarbonTracker (which also uses Key Biscayne data for those years) is less susceptible to such swings because many more data points (one order of magnitude) control the annual mean fluxes. In fact, one of our sensitivity tests demonstrates that removing half our data at random, or all of our tower sites, does not push the annual mean North American fluxes outside the range quoted in this work.

S6 Surface CO_2 residuals and bias

Similar to Figure 4 in the main text, a seasonal view of the residuals (Model-Observed) in CO_2 can be made to assess the bias of the optimized CO_2 field against the assimilated observations. This bias has to be close to zero in a well-designed assimilation experiment suggesting that the errors in the modeling framework are indeed random. Figure 4 shows the month-by-month residual statistics.

A seasonal cycle is discernible in the residuals indicating a seasonal bias in the performance of the system. The model produced too low mole fractions in winter (DJF, -0.42±2.15 ppm) , and too high in summer (JJA, 0.55±3.06) even after the assimilation. We speculate that this is due to the dynamical model that was applied (see Equation 1) which always pulls our analysis towards the prior fluxes from the CASA GFED model. If that model has a seasonal cycle that is too small [23] this would be reflected as the residuals in the figure. Further analysis of both the

Table 4: 2001-2005 annual average aggregated fluxes (mean± one standard deviation from covariance matrix) for TransCom regions from our system compared to similar estimates from Gurney *et al* [21], Baker *et al* [14], and Jacobson *et al* [18]. The time span of each of these studies is indicated in the table. A standard deviation of 999. indicates that this quantity was not calculated.

Region Name	This work 2000-2006	Baker *et al* [14] 1991-2000	Gurney *et al* [21] 1991-1996	Jacobson *et al* [18] 1995(*)
(1) North America Boreal	-0.16± 0.44	+0.14± 0.19	+0.20± 0.33	-0.13± 0.64
(2) North America Temperate	-0.50± 0.60	-1.11± 0.23	-0.89± 0.39	-0.93± 0.76
(3) South America Tropical	+0.02± 0.66	+1.07± 0.69	+0.74± 1.06	+3.07± 2.36
(4) South America temperate	+0.09± 0.81	-0.64± 0.51	-0.24± 0.88	-0.84± 1.72
(5) Northern Africa	-0.00± 0.52	+0.50± 0.50	+0.79± 1.01	+1.62± 2.95
(6) Southern Africa	+0.20± 0.59	-0.62± 0.48	-0.51± 0.84	-1.74± 2.57
(7) Eurasian Boreal	-0.58± 1.26	-0.33± 0.24	-0.36± 0.56	+0.04± 0.72
(8) Eurasian Temperate	-0.32± 0.63	-0.31± 0.25	-0.41± 0.81	-0.81± 1.20
(9) Tropical Asia	+0.08± 0.21	+0.29± 0.31	+0.27± 1.04	-0.50± 2.19
(10) Australia	-0.09± 0.33	-0.11± 0.12	-0.10± 0.22	+0.15± 0.35
(11) Europe	-0.25± 0.78	-0.97± 0.19	-0.96± 0.47	-1.05± 0.54
(12) North Pacific Temperate	-0.43± 0.25	-0.56± 0.14	-0.32± 0.31	-0.45± 0.08
(13) West Pacific Tropics	+0.12± 0.05	-0.11± 0.13	-0.21± 0.32	+0.12± 0.07
(14) East Pacific Tropics	+0.55± 0.19	+0.57± 0.14	+0.66± 0.33	+0.31± 0.03
(15) South Pacific Temperate	-0.29± 0.14	+0.09± 0.20	+0.51± 0.57	-0.50± 0.07
(16) Northern Ocean	-0.34± 0.13	-0.22± 0.09	-0.27± 0.19	-0.21± 0.07
(17) North Atlantic Temperate	-0.26± 0.14	-0.29± 0.13	-0.29± 0.33	-0.40± 0.06
(18) Atlantic Tropics	+0.11± 0.04	+0.09± 0.14	-0.10± 0.24	+0.20± 0.11
(19) South Atlantic Temperate	-0.16± 0.08	-0.21± 0.15	-0.05± 0.25	-0.24± 0.05
(20) Southern Ocean	-0.50± 0.29	-0.25± 0.11	-0.55± 0.37	-0.15± 0.07
(21) Indian Tropical	+0.17± 0.08	+0.24± 0.18	-0.33± 0.33	+0.13± 0.07
(22) South Indian Temperate	-0.45± 0.20	-0.41± 0.12	-0.39± 0.29	-0.52± 0.04
(23) Non Optimized	-0.02± 0.00	+0.00± 0.00	+0.00± 0.00	+0.00± 0.00
(24) Global Land	-1.51± 2.27	-2.09± 0.53	-1.46± 0.98	-1.12± 0.23
(25) Global Ocean	-1.48± 0.54	-1.06± 0.47	-1.34± 0.98	-1.71± 0.22
(26) Global Total	-3.01± 2.33	-3.15± 0.25	-2.81± 0.01	-2.83± 0.07
(27) North America Total	-0.65± 0.75	-0.97± 0.25	-0.69± 0.00	-1.06± 0.00
(28) NH Land	-1.80± 999.	-2.58± 999.	-2.42± 999.	-2.88± 999.
(29) Tropical Land	+0.09± 999.	+1.86± 999.	+1.80± 0.00	+4.19± 999.
(30) NH Total	-2.83± 999.	-3.65± 999.	-3.30± 999.	-3.94± 999.
(31) Tropical Total	+1.05± 999.	+2.65± 999.	+1.82± 999.	+4.95± 999.

(*) Estimate for post-industrial era scaled to 1995 fossil fuel contribution

CASA model and the CarbonTracker output, for instance against $^{13}CO_2$ could possibly inform us on future improvements.

Interestingly, the bias at the surface does not seem to correlate fully in size and magnitude with the FT bias, nor lead it in an obvious way. The sign of the bias is the same though in 9 out of 12 months, suggesting that the surface bias is at least partly responsible for the FT bias. Another part of the FT differences must be due to shortcomings in the modeled vertical transport as described also by Stephens *et al.* [22] and Yang *et al* [23].

Assessing the influence of these biases on the flux results in nearly impossible without creating an unbiased model first. The best we can therefore do is a simpler analysis of the bias magnitude in relation to other components of our model. For instance, we can consider the 0.27 ppm FT bias in summer to be a 'signal' in the free troposphere resulting from fluxes that still need to be adjusted. The influence of this 'signal' on the fluxes will depend on a) the sensitivity of the surface fluxes to the signal, b) the model's skill in reproducing the observations relative to the signal, and c) the magnitude of this signal relative to other signals pulling the fluxes. From additional analysis of our results for June/July/August 2004, we have determined that the uncertainty in the current set of summer fluxes causes an average spread of 2.1 ppm averaged over all aircraft observations. This suggests the 0.27 ppm mixing ratio signal to be nearly an order of magnitude smaller than the uncertainty in the random flux component we try to optimize and therefore unlikely to exert much influence on the results (point a). The error bars in Figure 4 in the main text reflect a similar conclusion derived differently: the bias is small compared to the model's skill in matching the observations in the first place and therefore the amount of flux information to be pulled from the signal is small (point b). Most importantly though, we can see that a false gradient of 0.27 ppm in the mixing ratio field is very small compared to the 'real' site-to-site gradients used by the inversion which are on average 7.1 ppm for this period (point c). Although it is not completely fair to compare a systematic bias to a random error component, these numbers suggest, at least to us, that the effect of a systematic bias of 0.27±2.7 ppm is likely to have been small. Nevertheless, we are investing substantial effort at the moment to reduce these biases in CarbonTracker by improving the first-guess surface fluxes and investigating transport biases.

S7 Column CO_2 at Park Falls, Wisconsin

The comparison of column CO_2 at Park Falls, Wisconsin was done by selecting all days where successful FTS observations were made between 9am and 10am local time (15-16 GMT). The available observations were averaged and compared to the CarbonTracker columns between 9:30am and 10:30 am local time for the same location. Before averaging, CarbonTracker mole fractions were pressure weighted and convolved with one fixed averaging kernel resembling the kernel shown in Figure 5b of Washenfelder *et al.* [24]. In addition, stratospheric CO_2 values in CarbonTracker were reset to a value of 368.0 ppm similar to the profile used in Washenfelder *et al.*. Generally, CarbonTracker stratospheric values are biased low as a result of low initialization and the slow exchange with the troposphere.

The time series of column average CO_2 is shown in Figure 5. Absolute differences between the two series shown are below 0.9 ppm for 68% of the points, but the mean offset is 0.5 ppm. This is partly due to relatively high summer values in CarbonTracker compared to the observations. This bias is similar to that seen in surface and free tropospheric data in that time of year. Note that part of the good agreement between the Park Falls FTS data and CarbonTracker might be due to

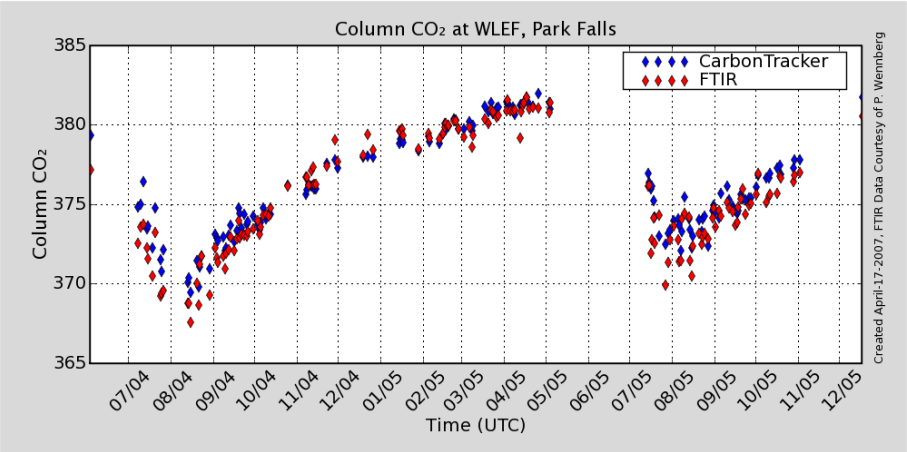

Figure 5: Comparison between column average CO_2 from observations and from our modeled CO_2 distribution for FTS observations at Parks Falls, Wisconsin, USA. The figure shows the general good agreement between modeled CO_2 and independent (non-assimilated) observations. The linear correlation coefficient is r=0.96. Units are ppm.

the assimilation of surface CO_2 observations from the tower at this location. Further tests of this hypothesis are planned for the near future.

References

[1] Global CO2 Emissions from Fossil-Fuel Burning, Cement Manufacture, and Gas Flaring: 1751-2003 (http://cdiac.ornl.gov/ftp/ndp030/global.1751_2003.ems)

[2] Olivier, J. G. J. & Berdowski, J. J. M. (2001) eds. Berdowski, J., Guicherit, R. & Heij, B. J. (A.A. Balkema Publishers/Swets and Zeitlinger Publishers, Lisse, The Netherlands), pp. 33–78.

[3] Blasing, T. J., Broniak, C. T. & Marland, G. (2005) *Tellus B* **57**, 107–115.

[4] Suntharalingam, P., Randerson, J. T., Krakauer, N., Logan, J. A. & Jacob, D. J. (2005) *Glob Biogeochem Cycles* **19**, GB4003.

[5] Van Der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S. & Arellano, J., A.F. (2006) *Atmospheric Chemistry and Physics* **6**, 3423–3441.

[6] Giglio, L., van der Werf, G. R., Randerson, J. T., Collatz, G. J. & Kasibhatla, P. (2006) *Atmospheric Chemistry and Physics* **6**, 957–974.

[7] Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., *et al.* (2002) *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **49**, 1601–1622.

[8] Kettle, H. & Merchant, C. J. (2005) *Atmospheric Chemistry and Physics* **5**, 1459–1466.

[9] Olsen, S. C. & Randerson, J. T. (2004) *Journal of Geophysical Research-Atmospheres* **109**,

[10] Kaminski, T., Heimann, M. & Giering, R. (1999) *Journal of Geophysical Research-Atmospheres* **104**, 18555-18581.

[11] Olson, J. S., Watts, J. A. & Allison, L. J. (1985) *NDP-017*

[12] Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L. & Tans, P. P. (2005) *Journal of Geophysical Research-Atmospheres* **110**, doi:10.1029/2005JD006157.

[13] Houtekamer, P. L. & Mitchell, H. L. (1998) *Monthly Weather Review* **126**, 796–811.

[14] Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., *et al.* (2006) *Global Biogeochemical Cycles Global Biogeochem. Cycles* **20**,

[15] Peylin, P., Baker, D., Sarmiento, J., Ciais, P. & Bousquet, P. (2002) *Journal of Geophysical Research-Atmospheres* **107**,

[16] Xiao,X (2007) *Journal of geophysical research* **112**, D07303.

[17] Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Youing, L. & Merchant, J. W. (2000) *Int. J. Remote Sensing* **21**, 1303–1330.

[18] Jacobson, A. R., J. L. Sarmiento, M. Gloor, N. Gruber, S. E. Mikaloff Fletcher & Modelers, T. C. (2006) *submitted to Glob. Biogeochemical Cycles*

[19] Rödenbeck, C., Houweling, S., Gloor, M. & Heimann, M. (2003) *Atmospheric Chemistry and Physics* **3**, 1919–1964.

[20] Andres, R. J., Marland, G., Fung, I. & Matthews, E. (1996) *Global Biogeochemical Cycles* **10**, 419–429.

[21] Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., *et al.* (2004) *Global Biogeochemical Cycles* **18**,

[22] Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., *et al.* (2007) *Science* **316**, 1732-1735

[23] Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., Sweeney, C. & Wennberg, P. O. (2007) *Geophys. Res. Lett.* **34**, doi:10.1029/2007GL029742

[24] Washenfelder, R. A., Toon, G. C., Blavier, J. F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M. & Daube, B. C. (2006) *Journal of Geophysical Research Atmospheres* **111**, D22305.