Impact of CO2 measurement bias on CarbonTracker surface flux estimates

Masarie, K. A.; PéTron, G.; Andrews, A.; Bruhwiler, L.; Conway, T. J.; Jacobson, A. R.; Miller, J. B.; Tans, P. P.; Worthy, D. E.; Peters, W.

Published in:
Journal of Geophysical Research

DOI:
10.1029/2011JD016270

Publication date:
2011

Impact of CO_2 measurement bias on CarbonTracker surface flux estimates

K. A. Masarie,[1] G. Pétron,[1,2] A. Andrews,[1] L. Bruhwiler,[1] T. J. Conway,[1] A. R. Jacobson,[1,2] J. B. Miller,[1,2] P. P. Tans,[1] D. E. Worthy,[3] and W. Peters[4]

Received 19 May 2011; revised 30 June 2011; accepted 8 July 2011; published 9 September 2011.

[1] For over 20 years, atmospheric measurements of CO_2 dry air mole fractions have been used to derive estimates of CO_2 surface fluxes. Historically, only a few research laboratories made these measurements. Today, many laboratories are making CO_2 observations using a variety of analysis techniques and, in some instances, using different calibration scales. As a result, the risk of biases in individual CO_2 mole fraction records, or even in complete monitoring networks, has increased over the last decades. Ongoing experiments comparing independent, well-calibrated measurements of atmospheric CO_2 show that biases can and do exist between measurement records. Biases in measurements create artificial spatial and temporal CO_2 gradients, which are then interpreted by an inversion system, leading to erroneous flux estimates. Here we evaluate the impact of a constant bias introduced into the National Oceanic and Atmospheric Administration (NOAA) quasi-continuous measurement record at the Park Falls, Wisconsin (LEF), tall tower site on CarbonTracker flux estimates. We derive a linear relationship between the magnitude of the introduced bias at LEF and the CarbonTracker surface flux responses. Temperate North American net flux estimates are most sensitive to a bias at LEF in our CarbonTracker inversion, and its linear response rate is 68 Tg C yr^{-1} (~10% of the estimated North American annual terrestrial uptake) for every 1 ppm of bias in the LEF record. This sensitivity increases when (1) measurement biases approached assumed model errors and (2) fewer other measurement records are available to anchor the flux estimates despite the presence of bias in one record. Flux estimate errors are also calculated beyond North America. For example, biospheric uptake in Europe and boreal Eurasia combined increases by 25 Tg C yr^{-1} per ppm CO_2 to partially compensate for changes in the North American flux totals. These results illustrate the importance of well-calibrated, high-precision CO_2 dry air mole fraction measurements, as well as the value of an effective strategy for detecting bias in measurements. This study stresses the need for a monitoring network with the necessary density to anchor regional, continental, and hemispheric fluxes more tightly and to lessen the impact of potentially undetected biases in observational networks operated by different national and international research programs.

Citation: Masarie, K. A., G. Pétron, A. Andrews, L. Bruhwiler, T. J. Conway, A. R. Jacobson, J. B. Miller, P. P. Tans, D. E. Worthy, and W. Peters (2011), Impact of CO_2 measurement bias on CarbonTracker surface flux estimates, *J. Geophys. Res.*, *116*, D17305, doi:10.1029/2011JD016270.

1. Introduction

[2] Over the last decade, efforts to independently estimate carbon balances on regional and national scales have led to an increase in CO_2 measurements made by different laboratories using new experimental techniques and sampling strategies. Prior to 2000, the number of research laboratories making long-term measurements of atmospheric CO_2 was small and focused primarily on understanding global and hemispheric flux patterns [e.g., *Conway et al.*, 1994]. Detection of CO_2 in air was primarily by nondispersive infrared (NDIR) absorption or gas chromatography equipped with a methanizer and flame ionization detector. Today, many national agencies, universities, and multinational projects are using an increasing variety of observational techniques to make ongoing long-term and campaign-based CO_2 observations. Two relatively new

[1]Global Monitoring Division, Earth System Research Laboratory, NOAA, Boulder, Colorado, USA.

[2]Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA.

[3]Environment Canada, Toronto, Canada.

[4]Department of Meteorology and Air Quality, Wageningen University, Wageningen, Netherlands.

0148-0227/11/2011JD016270

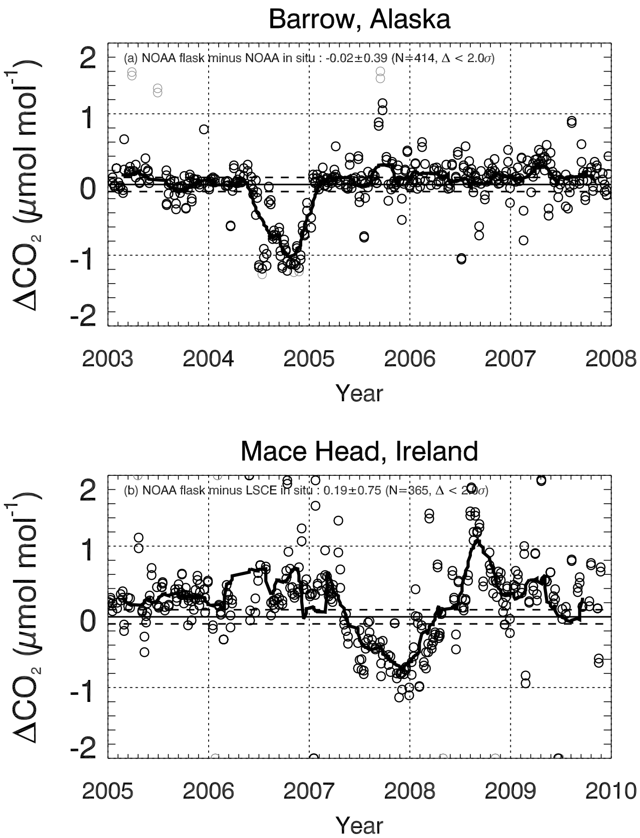

Figure 1. (top) Weekly discrete samples collected at Barrow, Alaska, and analyzed at NOAA ESRL in Boulder are compared with hourly averaged values derived from NOAA quasi-continuous measurements made in situ (flask minus in situ) for 2003–2007. (bottom) Weekly discrete samples collected at Mace Head, Ireland, and analyzed at NOAA in Boulder are compared with hourly averaged values derived from LSCE quasi-continuous measurements made in situ (flask minus in situ) for 2005–2009. The horizontal dashed lines at ±0.1 ppm identify the WMO recommended target level for interlaboratory compatibility. The figure shows the great challenge of meeting the WMO targets over longer time periods and that biases across different instruments and different laboratories are not constant over time.

programs include (1) the joint European Integrated Carbon Observing System (ICOS) project, which will operate a cooperative European network of stations equipped with cavity ring-down spectroscopy (CRDS) detectors to make long-term high-precision quasi-continuous CO_2 measurements (see http://www.icos-infrastructure.eu), and (2) the Japan multiagency program making routine high-precision quasi-continuous NDIR CO_2 measurements aboard commercial passenger aircraft [*Machida et al.*, 2008].

[3] In addition to measurements using closed-path measurement techniques (NDIR or CRDS), a new category of space- and ground-based open-path spectroscopic techniques is emerging. These measurements are especially prone to biases in the CO_2 mole fraction because the measurement path cannot be filled with air directly traceable to the World Meteorological Organization (WMO) CO_2 mole fraction scale [*Zhao and Tans*, 2006; *Tans et al.*, 2011]. The Total Carbon Column Observing Network (TCCON), established in 2009, is a multinational collaboration coordinated by the California Institute of Technology and includes 15 sites worldwide. TCCON uses ground-based Fourier transform spectrometers (FTS) to measure solar spectra, which are used to retrieve column-averaged abundances of CO_2 and other trace constituents [*Wunch et al.*, 2010]. Also in 2009, the Ministry of the Environment, the National Institute for Environmental Studies, and the Japan Aerospace Exploration Agency successfully deployed the Greenhouse Gases Observing Satellite, the first satellite mission specifically dedicated to providing space-based retrievals of column CO_2 [*Morino et al.*, 2011].

[4] The introduction of new measurement programs is a critical step toward addressing the severe lack of atmospheric CO_2 observations, which has been a primary obstacle to producing reliable estimates of net CO_2 exchange at the Earth's surface using atmospheric inversion techniques. More observations, however, are only useful to this effort if they can be merged with existing measurements without introducing significant bias into calculated surface fluxes. Bias in closed-path measurement records may occur when undetected problems in gas handling and detection, data processing, and propagation of the calibration scale introduce systematic errors. For open-path systems, biases are more likely because the CO_2 retrievals cannot be calibrated, as we cannot control what is in the optical path. They can only be compared to each other and to infrequent well-calibrated in situ chemical measurements of the part of the atmospheric column that can be reached by aircraft or balloon. The radiative measurement can be calibrated relative to radiation standards, but the CO_2 retrieval is a data assimilation using a radiative transfer model that includes atmospheric properties and optical parameters, some of which may not be characterized well enough.

[5] Our best strategy for detecting bias in measurements has been to make routine comparisons of independent and colocated atmospheric measurements. Comparison experiments can be within a single laboratory or between laboratories. For example, at the National Oceanic and Atmospheric Administration (NOAA) baseline observatories and tall tower sites, we routinely compare quasi-continuous in situ measurements with air samples collected weekly (or daily) in the field and analyzed in Boulder, Colorado. These comparisons directly evaluate the performance of different sampling strategies and analytical techniques used throughout the NOAA observing network and are critical to evaluating the quality of our measurements. Figure 1 (top) shows results from an ongoing CO_2 comparison of weekly air samples collected in glass flasks at Barrow, Alaska, and measured at the NOAA Earth System Research Laboratory (ESRL) in Boulder with hourly averaged values derived from NOAA quasi-continuous CO_2 measurements made in situ. This within-laboratory comparison shows a mean difference between the flask and in situ records of -0.02 ± 0.39 ppm (where uncertainty is reported at the 68% confidence level or 1σ and 1 ppm $\equiv$ 1 μmol mol^{-1}) averaged over 2003–2007. On shorter time scales, however, we calculate statistically significant differences that can persist for several months. In 2004, for example, we observed a 1 ppm bias between the two records

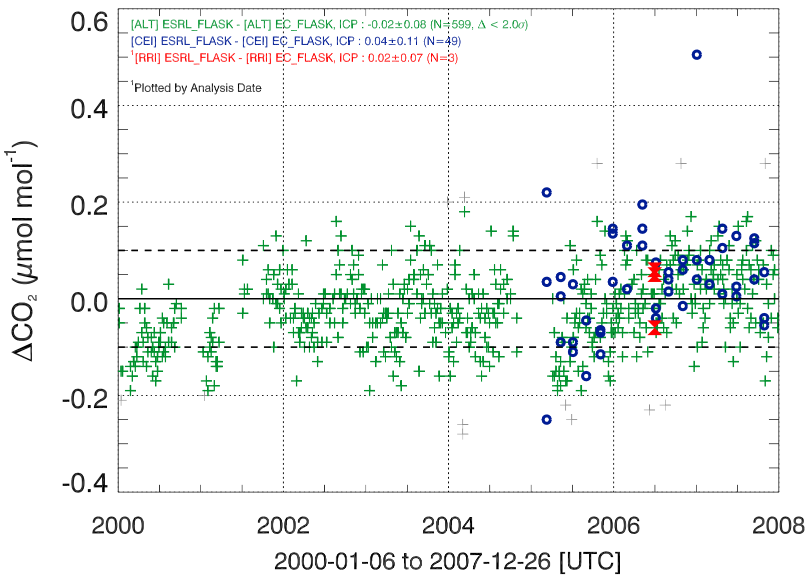

Figure 2. Comparison results of independent CO_2 measurements made by NOAA and Environment Canada (EC). Measurement differences from weekly air samples collected in NOAA and EC glass flasks and analyzed in their respective labs are shown as green pluses. Measurement differences from bimonthly air samples collected in NOAA and EC flasks from air in a high-pressure cylinder are shown as open blue circles. Measurement differences from the most recent WMO round robin experiment are shown in red. The horizontal dashed lines at ±0.1 ppm identify the WMO recommended target level for interlaboratory compatibility.

lasting for nearly 6 months. This discrepancy between the two records remained unexplained until only recently when it was discovered that the CO_2 mole fraction in one of three reference gas cylinders used to calibrate the response of the NDIR detector at Barrow had drifted in a very atypical manner. This finding accounts for ~0.6 ppm of the observed bias. Figure 1 (bottom) shows results from an interlaboratory comparison between the Laboratoire des Sciences du Climat et l'Environnement (LSCE) in France and NOAA based on measurements from Mace Head, Ireland. NOAA measurements of weekly air samples are compared with hourly averaged values derived from quasi-continuous CO_2 measurements made in situ by LSCE. The mean difference (flask minus in situ) is 0.18 ± 0.75 ppm over 2005–2009. Again, over shorter averaging periods, we observe a mean bias of −0.51 ± 0.33 ppm lasting several months in 2007–2008 and then changing sign (+0.45 ± 0.55 ppm) in mid-2008 and persisting for several months. Results from complementary comparisons suggest the problem may be due, in part, to problems with the in situ measurements, but this has not yet been confirmed. These examples illustrate features common to many comparison experiments. Despite careful attention, biases can and in fact do exist. Measurement difference distributions can vary over time with large biases sometimes lasting several months, and, while not shown, differences can change retroactively when data are reprocessed or transferred to new calibration scales.

[6] The internationally recognized target levels for compatibility between independent observations of atmospheric CO_2 are ±0.1 ppm and ±0.05 ppm in the Northern and Southern Hemispheres, respectively [*World Meteorological Organization*, 2009]. While these targets have proven to be a difficult technical challenge, they are attainable as demonstrated by several direct atmospheric comparison experiments [*Masarie et al.*, 2001, 2009]. Figure 2 shows results from weekly, bimonthly, and multiyear same-air comparisons of measurements from Environment Canada (EC) and NOAA. All results are coherent and suggest that observations made by the two labs are compatible to well within the ±0.1 ppm target level. Table 1 summarizes CO_2 mole fraction comparison results for a subset of ongoing NOAA, EC, Commonwealth Scientific and Industrial Research Organization (CSIRO), and National Center for Atmospheric Research (NCAR) comparison experiments relevant to CarbonTracker.

[7] The impact of measurement bias on flux estimates derived from an inversion system will depend on the spatial and temporal extents of the biases and network density and the balance between prior flux and data uncertainty in the inversion setup. Errors in model transport, flux optimization, prescribed and first-guess model input data, and atmospheric observations all contribute to uncertainty in the estimated fluxes. Many of the errors in the inversion system are formally estimated and used in the calculations, but they are often poorly known and the propagation assumes random error distribution. Systematic errors are sometimes included too [*Dee and da Silva*, 1999; *Engelen et al.*, 2009] but are often unknown or not characterized at all, leading to biased CO_2 surface flux estimates.

[8] *Rödenbeck et al.* [2006] examined the impact reported experimental errors have on CO_2 fluxes estimated using a global multiyear CO_2 inversion system. They used observations from the NOAA cooperative air sampling network and the CSIRO observing network. NOAA and

Table 1. Summary of a Subset of NOAA, EC, CSIRO, and NCAR Comparison Experiments for the Period 2000–2007[a]

Year	ALT (EC)[b]	CGO (CSIRO)[b]	LEF (NOAA)[c]	MLO (NOAA)[d]	NWR (NCAR)[d]
2000	0.05 ± 0.11 (119)	0.16 ± 0.16 (88)		0.13 ± 0.14 (81)	
2001	−0.03 ± 0.14 (102)	0.14 ± 0.16 (70)		0.03 ± 0.14 (86)	
2002	−0.04 ± 0.11 (141)	0.03 ± 0.14 (70)		0.03 ± 0.12 (88)	
2003	−0.04 ± 0.12 (154)	0.12 ± 0.14 (63)		0.04 ± 0.09 (80)	
2004	−0.01 ± 0.13 (113)	0.10 ± 0.15 (75)		0.00 ± 0.13 (86)	
2005	−0.00 ± 0.13 (132)	0.13 ± 0.15 (45)		0.06 ± 0.13 (83)	
2006	−0.07 ± 0.14 (161)	0.13 ± 0.14 (45)	−0.00 ± 0.46 (72)	0.10 ± 0.12 (77)	−0.03 ± 0.26 (96)
2007	−0.09 ± 0.13 (168)	0.07 ± 0.12 (40)	0.04 ± 0.28 (87)	0.15 ± 0.13 (78)	0.09 ± 0.26 (86)

[a]Units for are all ppm (ppm ≡ μmol mol^{-1}).
[b]Comparison of weekly colocated flask measurements at Alert, Nunavut, Canada (ALT) and Cape Grim, Tasmania, Australia (CGO).
[c]Comparison of weekly colocated flask and quasi-continuous measurements at the 396 magl inlet at LEF.
[d]Comparison of weekly colocated flask and quasi-continuous measurements at Mauna Loa, Hawaii (MLO) and Niwot Ridge, Colorado (NWR).

CSIRO have been comparing measurements of the same air from weekly samples collected at Cape Grim, Tasmania, since 1992. This long-term comparison shows that CSIRO and NOAA measurements based on this experiment have been compatible to within 0.2 ppm (Table 1) [*Masarie et al.*, 2001]. The NOAA and CSIRO data used by Rödenbeck *et al.* were primarily measurements of large, well-mixed air masses representative for scales of thousands of kilometers. In one case study, Rödenbeck *et al.* directly inverted a smoothed representation of observed differences at five remote observatories to estimate errors in net surface fluxes. They concluded that observed measurement biases within and between the CSIRO and NOAA networks have a small impact on their inversion results compared to other errors within the inversion system.

[9] Since the *Rödenbeck et al.* [2006] study, the CO_2 observing network has expanded further, especially in North America and Europe, and especially using quasi-continuous analyzers at tall tower sampling sites [e.g., *Vermeulen et al.*, 2011; *Winderlich et al.*, 2010; *Popa et al.*, 2010]. Large variability, large data volume, and shortcomings of the transport models are sometimes invoked to justify loosening the requirements on observational precision and accuracy for such sites. But new data inversion systems such as CarbonTracker [*Peters et al.*, 2007, 2010] are now using CO_2 data from these platforms to constrain regional exchange estimates. This motivates us to reexamine the role of biases specifically when nonsmoothed, daily data from continental locations are used rather than slowly varying mixing ratios from sites in the remote background atmosphere.

[10] In this study, we evaluate the sensitivity of flux estimates by CarbonTracker (described in section 2.2) to a constant-in-time bias introduced into a single quasi-continuous measurement record from the NOAA tall tower at Park Falls, Wisconsin (LEF). CO_2 measurements at LEF are influenced by surrounding forests and by boreal forests and croplands farther afield as well as the large-scale Northern Hemisphere CO_2 spatial patterns [*Bakwin et al.*, 2004]. We describe our test simulations (section 2.3) and present results of how CarbonTracker flux estimates change with introduced bias at a single site in North America (section 3). We focus our analysis on three questions raised when dealing with a biased CO_2 measurement record: (1) What is the impact on fluxes that are directly constrained by the biased site? (2) How do biases propagate beyond the area directly constrained? (3) How are the results influenced by the presence of other sites in a dense regional network?

2. Methods

2.1. Site Description

[11] In this study we focus on the effect of measurement bias at one site, the Wisconsin tall tower (LEF). The tower (45.95°N, 90.27°W), 472 m above sea level (masl)) is located in a heavily forested zone of low relief [*Bakwin et al.*, 1995]. Canadian boreal forests are to the north and northeast of the site and agricultural lands of the U.S. Midwest to the south and southwest. Quasi-continuous measurements of CO_2 are made at three levels ranging from 30 to 396 m above ground level (magl). CarbonTracker uses mole fractions averaged between 12:00 and 16:00 local standard time from the highest intake. Cluster analyses of back trajectories derived from the Stochastic Time-Inverted Lagrangian Transport Model (STILT) driven by meteorological fields from the Weather Research and Forecasting Model (WRF) suggest that ~50% of the trajectories arriving at the LEF tower are from the north and northwest and ~42% from the south.

2.2. CarbonTracker

[12] CarbonTracker is a data inversion system designed to calculate net surface land and ocean CO_2 fluxes that are consistent with high-precision atmospheric CO_2 mole fraction measurements from a global well-calibrated measurement network [*Peters et al.*, 2007]. CarbonTracker creates weekly forecasts of CO_2 mole fractions on a global grid using an off-line atmospheric transport model and four surface flux modules: (1) prescribed anthropogenic fossil fuel emissions, (2) prescribed fire emissions [*van der Werf et al.*, 2006], (3) first-guess (a priori) terrestrial net ecosystem exchange (NEE) [*van der Werf et al.*, 2006], and (4) a priori net ocean surface fluxes [*Jacobson et al.*, 2007]. The resulting four-dimensional (4-D) atmospheric CO_2 distribution (x,y,z,t) is then interpolated for the times and locations where atmospheric observations exist. Using an ensemble Kalman filter optimization scheme with a 5-week lag, the initial land and ocean net surface fluxes are adjusted up or down using a set of weekly and regional flux scaling factors to minimize differences between the forecasted and observed CO_2 mole fractions. From *Peters et al.*

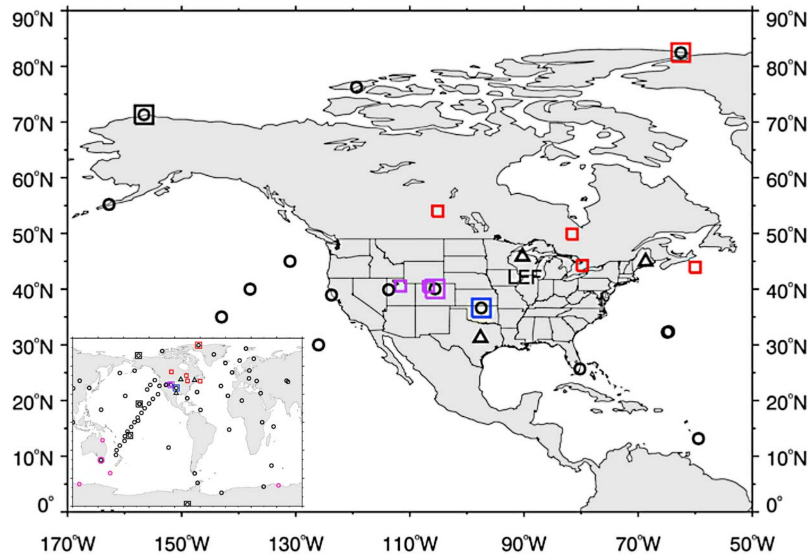

Figure 3. CarbonTracker 2008 Network Distribution includes sites from NOAA (black), EC (red), CSIRO (magenta), NCAR (purple), and LBNL (blue). The NOAA tall tower in Wisconsin (LEF) is identified.

[2007], the optimized surface flux estimates as a function of time, F(x,y,t), have the form

$$F(x, y, t) = \lambda_r \cdot F_{bio}(x, y, t) + \lambda_r \cdot F_{oce}(x, y, t) + F_{ff}(x, y, t) + F_{fire}(x, y, t), \quad (1)$$

where F_{bio} and F_{oce} are the a priori land biosphere and ocean fluxes, F_{ff} and F_{fire} are prescribed fossil fuel and fire emissions, and λ_r is a set of weekly scaling factors applied to land and ocean surface fluxes at the ocean basin and ecoregion scale. The adjusted fluxes are then assimilated to create a 4-D CO_2 distribution that is optimally consistent with the observations. CarbonTracker uses the two-way nested TM5 transport model [*Krol et al.*, 2005 *Huijnen et al.*, 2010] run at a global 6° x 4° resolution with nested regional grids over North America (3° × 2°) and the United States (1° × 1°) [*Peters et al.*, 2004]. CarbonTracker has been updated annually since its first release in 2007. In this work, we use CarbonTracker 2008 (CT2008), which provides optimized fluxes for the period 2000–2007.

[13] The observing sites used in CT2008 are shown in Figure 3. CT2008 includes measurements from ~18,000 weekly air samples from across the world (circles), ~20,000 afternoon-averaged values derived from quasi-continuous CO_2 measurements at sites located primarily in North America (squares), and ~9,000 afternoon-averaged values from tall towers at three locations within the continent (triangles). At many of the quasi-continuous sampling sites, we construct an afternoon average mole fraction for each day from the time series, recognizing that our atmospheric transport model does not capture well the continental nighttime strong near-surface vertical gradients. At mountaintop sites (Mauna Loa, Hawaii (MLO), Niwot Ridge, Colorado (NWR), and Storm Peak, Colorado (SPL)), we use an average of nighttime hours as this tends to be the most stable time period, avoiding convection-driven upslope flow conditions that can transport local vegetative and anthropogenic signals to the measurement site. At each time step, the sites for which data are available vary depending on successful sampling and analysis and each site's sampling frequency. Data with known experimental problems or thought to be influenced by local sources or sinks are excluded from this study.

[14] CT2008 includes observations from five laboratories: NOAA, CSIRO, EC, NCAR and Lawrence Berkeley National Laboratory (LBNL). While not evident from Figure 3, the current CarbonTracker network is a subset of a much larger network of observations made by more than 50 laboratories worldwide [e.g., *World Data Centre for Greenhouse Gases*, 2011; *GLOBALVIEW-CO_2*, 2010]. We include measurements from these five laboratories because they all report data on the WMO CO_2 mole fraction scale, and results from ongoing direct comparisons of atmospheric measurements at colocated sampling sites spanning the time period 2000–2007 suggest the measurements are generally compatible to within a mean difference of 0.3 ppm. This estimate of network measurement uncertainty contributes a minor part to the CarbonTracker model-data mismatch (MDM) term, which weights each assimilated observation based on the expected skill with which the model can reproduce it.

2.3. Sensitivity Experiments

[15] We compare CarbonTracker simulations (test runs), where we have introduced bias into the observations, to CarbonTracker control runs (Table 2). For all but one test run, we introduce a constant bias for a fixed period of time to the LEF data set; all other data remain unchanged. Test runs differ from each other by the magnitude of the introduced offset and the time period for which the bias is introduced. The CarbonTracker inversion system is run for a 16-month period spanning the calendar year in which the bias is introduced. A simulation in which bias is added to

Table 2. 2004 and 2007 CarbonTracker Sensitivity Test Simulations

	Description
2004A	Test runs include a constant offset added to all available LEF 2004 afternoon-averaged values. Offsets range from −1 to +5 ppm (−1, +0.25, 0.5, 0.75, 1, 1.5, 3, 5).
2004B	Runs are identical to 2004A except the introduced offsets are large compared to the assumed Gaussian model uncertainty (3 ppm) and range from 8.5 to 15 ppm (8.5, 10, 12, 15).
2004C	Test runs include a constant offset added to all available LEF 2004 afternoon-averaged values as in experiment 2004A. However, both the control run (referred to as CT2008*) and the test runs in 2004C exclude five EC quasi-continuous records located near or in boreal North America (ALT_06C0, CDL_06C3, EGB_06C3, FRD_06C3, and SBL_06C3). Four test runs were performed with offsets ranging from −1.0 to +3.0 ppm (−1, +1, 1.5, 3).
2004D	A single test run where we have removed all 2004 LEF data, i.e., introduced a 1year gap. No bias is introduced.
2007A	Runs are identical to 2004A except the offsets are introduced to all available 2007 (not 2004) LEF afternoon-averaged values. Offsets range from −1 to +5 ppm (−1, +0.5, 1, 3, 5).
2007B	This run is identical to 2007A runs except the introduced offset is large (12 ppm) compared to the assumed Gaussian model uncertainty (3 ppm).
2007C	This run introduces measurement gaps in the LEF 2007 data corresponding to the actual gaps existing in the LEF 2004 data. No offset is introduced.

2004 LEF measurements, for example, begins in mid-October 2003 and ends in February 2005. Test runs are initialized using results from the control run (CT2008).

[16] The 2004A–2004D simulations use 2004 afternoon-averaged values. This time period was selected because 2004 was thought to be a typical year with no significant El Niño-La Niña events. Further, 2004 is well into the CarbonTracker inversion period (2000–2007) and free from "spin-up" flux adjustments that occur during the first 18 months of the model run. We also performed a set of simulations using 2007 afternoon-averaged values from LEF. The 2007A–2007B simulations tested the robustness of the results from the 2004 runs. We chose to separately assess the regime where introduced offsets are small (ranging from −1 to +5 ppm) relative to the assumed Gaussian model uncertainties ($1\sigma = 3$ ppm), from the regime where the offsets are large (exceeding $3 \times 1\sigma$). In the latter regime (2004B and 2007B), CarbonTracker has the freedom to reject individual observations, which can introduce a nonlinear response in estimated fluxes. Finally, we tested our results with and without the presence of a set of five sites in the vicinity of the LEF tower (2004C), which can potentially help to anchor flux estimates even in the presence of the introduced bias.

3. Results

[17] We will first present our results for the experiments with smaller offsets and analyze their impact on inferred continental-scale annual mean fluxes (section 3.1.1), the seasonal cycle of continental fluxes (section 3.1.2), and on fluxes at the subcontinental level (section 3.1.3). After a brief discussion of the results for large offsets (section 3.2) and the impact of data gaps on CarbonTracker flux estimates (section 3.3), we will assess the effect of these local biases on global flux estimates (section 3.4).

3.1. Sensitivity to Bias in the Range −1 to +5 ppm

[18] The 2004A and 2007A simulations were designed to assess CarbonTracker flux responses to measurement bias introduced into observations from a single site in North America. For each simulation an offset was added to each afternoon-averaged value available for a specified year. The introduced bias ranged from −1 to +5 ppm. While results from many of our comparison experiments suggest measurement compatibility on the order of 0.2–0.3 ppm when averaged over multiple years, we do observe biases as large as 1–2 ppm lasting several months. We focus our discussion primarily on temperate and boreal North America, the two flux regions most affected by the introduced bias at LEF.

3.1.1. Annual Transcom-Region Averages

[19] Figures 4 and 5 show total annual flux differences (test run minus control run) for optimized fluxes in temperate and boreal North America. When the prescribed bias is between −1 and +5 ppm, a linear relationship exists between the magnitude of the introduced bias and the change in the total annual flux estimates for all regions. For the 2004A simulations, the temperate and boreal North America flux responses are 68 ± 1 and 10 ± 1 Tg C yr^{-1} (Tg C ≡ 10^{12} g carbon) per ppm CO_2 bias, respectively (solid magenta circles). The linear response is a result of the implementation of the ensemble Kalman filter, which is an approximation of the analytical solution to the inverse problem. While the linear response to introduced bias is a robust feature, the slope of the flux response depends on the inversion configuration. We evaluate this dependency further using the 2004C simulations. 2004C simulations are similar to 2004A except we use a CT2008* inversion configuration. CT2008* is identical to CT2008 except that the five EC quasi-continuous records located near or in boreal North America (ALT_06C0, CDL_06C3, EGB_06C3, FRD_06C3, and SBL_06C3) are excluded for the entire 8-year inversion period. Four test runs were performed with offsets ranging from −1.0 to +3.0 ppm. Flux results from the 2004C simulation are compared with CT2008* and shown in Figures 4 and 5 (open green circles). Results show a stronger sensitivity to the introduced bias in both temperate and boreal North America, with the greatest change occurring in the boreal region where the response is ~3.5 times that of 2004A. The general explanation of larger flux bias is a result of the temperate and boreal North American regions having fewer data constraints in CT2008*; the five EC sites removed in this case (2004C) are especially sensitive to fluxes from the boreal region, leaving the biased LEF data as the dominant remaining constraint. Indeed, back

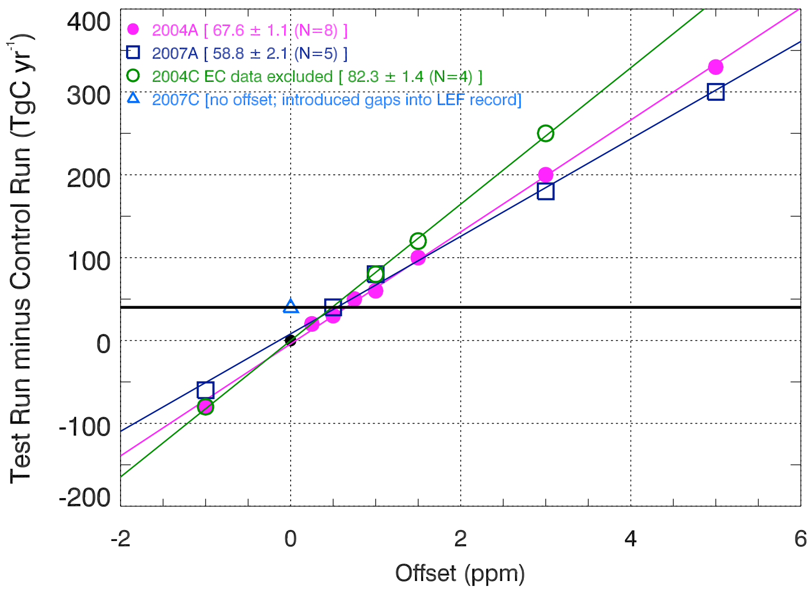

Figure 4. Change in CarbonTracker total annual net CO_2 exchange estimates (test minus control) for temperate North America as a function of introduced measurement bias at LEF. Results are from the 2004A, 2007A, 2004C, and 2007C simulations. The horizontal black line indicates the decrease in estimated uptake when 2004 LEF measurements are excluded. The estimated temperate North American carbon flux changes by 68 Tg C yr^{-1} for every 1 ppm of bias introduced at LEF.

trajectory analysis shows that airflow to LEF is from the north and northwest approximately 50% of the time. Without the EC observations, CarbonTracker can more freely adjust fluxes in the boreal region in response to the introduced bias.

[20] While CarbonTracker net biospheric flux errors increase linearly with introduced bias at LEF, derived CarbonTracker CO_2 mole fractions change in a more complicated way. Figure 6 shows the mean and standard deviation of the residuals (CarbonTracker mole fractions minus observations) for the control run (CT2008) and the 2004A (+1.5 ppm) test run. The difference is ~1 ppm with a slight CarbonTracker overestimate in the control run and an underestimate in the test run. The figure suggests that the biased observations pull the model results toward the new mole fractions successfully, but there is increasing inconsistency with the prior fluxes and other sites, as demonstrated from the decreased overall performance of the inversion. Interestingly, the biased observations show

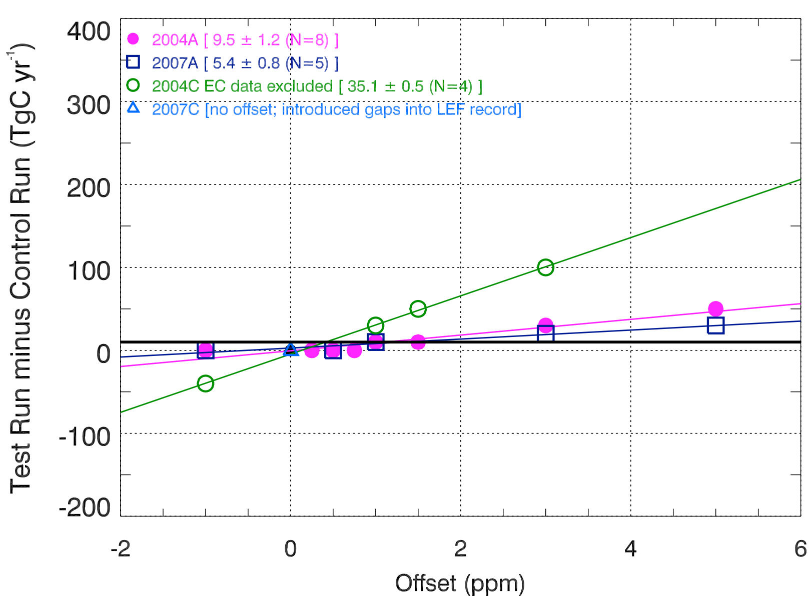

Figure 5. Same as Figure 4 but for boreal North America. Note that the change in flux is in the same direction as over temperate North America.

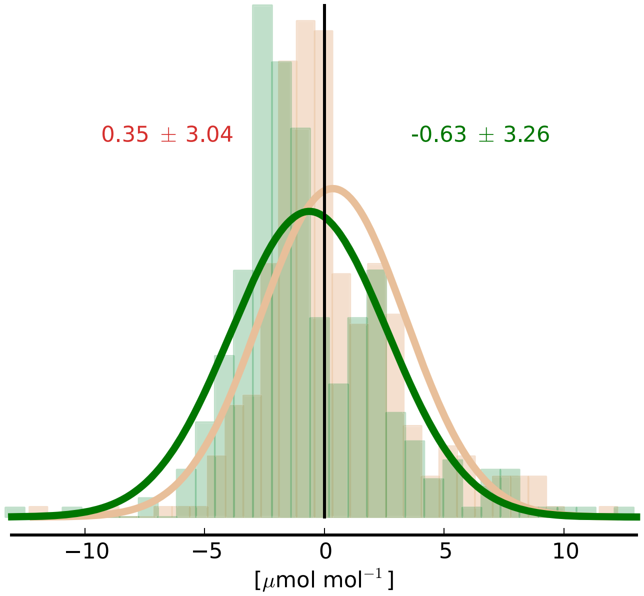

Figure 6. Histogram of 2004 LEF residuals (CT2008 minus observations) is shown in tan. The histogram of 2004 LEF residuals (CT2008 minus biased observations) from the 2004A (+1.5 ppm) simulation is overlaid in green. The mean and standard deviation of the residuals are reported in the same colors. Figure 6 shows that assimilating biased observations deteriorates the performance of the inversion system due to anchoring of the fluxes by other (unbiased) sites and by prior fluxes.

marked improvement in the mean of the residuals in summertime compared to CT2008, which is known to overestimate mixing ratios in summer over North America. Biased observations thus fit the biased model better in that period, at the expense of much worse agreement in wintertime. This suggests that the interaction between measurement bias and the presence of anchor sites and prior fluxes is complex, and an inversion system is often not sufficient to detect biases in measurement records. We do note, however, that *Law et al.* [2003] and *Miller et al.* [2007] successfully detected (self introduced) observational biases through a modeling system, so our result might be specific to this work.

3.1.2. Seasonal Averages

[21] We have applied a constant bias to the LEF data for an entire year. Figure 7 shows that the resulting CarbonTracker net flux response to a constant bias varies seasonally. The mean seasonal flux difference is strongest during June, July, and April (JJA) and December, January, and February (DJF), when the magnitude of a priori NEE that CarbonTracker can scale up or down is large; conversely, it is weakest during the transition seasons (March, April, and May (MAM) and September, October, and November (SON)) when the magnitude of NEE is small. Since the only difference between the simulations and the control run is the introduced constant offset, the seasonal response is due to the seasonally varying interaction between the offset and the model transport and prior fluxes.

3.1.3. Annual Ecoregion Averages

[22] The forest-field ecoregion [*Olson et al.*, 1985], located primarily in the southeastern United States, shows the strongest sensitivity to introduced measurement bias at LEF. Figure 8 show a map of 1° × 1° seasonal flux differences for the 2004A (+1.5 ppm) test run. CarbonTracker responds to the apparent increase in observed CO_2 at LEF by decreasing the NEE (uptake) in both boreal North America (18 Tg C) and the southeastern United States (16 Tg C) by nearly equal amounts. Boreal North American fluxes are constrained by other sites in the CarbonTracker Observing Network, specifically the quasi-continuous observations made by EC. The southeast, on the other hand,

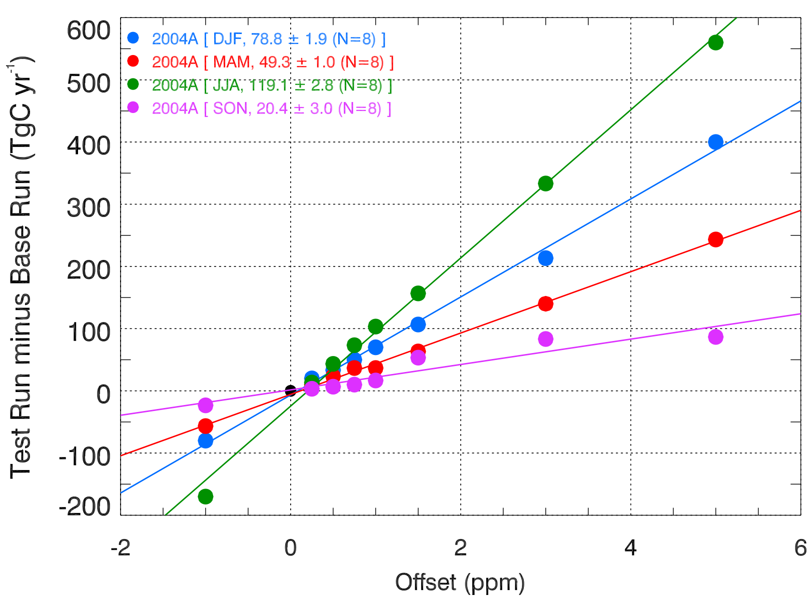

Figure 7. Seasonal change in CarbonTracker net CO_2 exchange estimates (test minus control) for temperate North America as a function of introduced measurement bias at LEF. Results are from the 2004A simulations. The response to a constant-in-time +1.5 ppm bias is clearly a function of season due to changing transport and a-priori flux patterns.

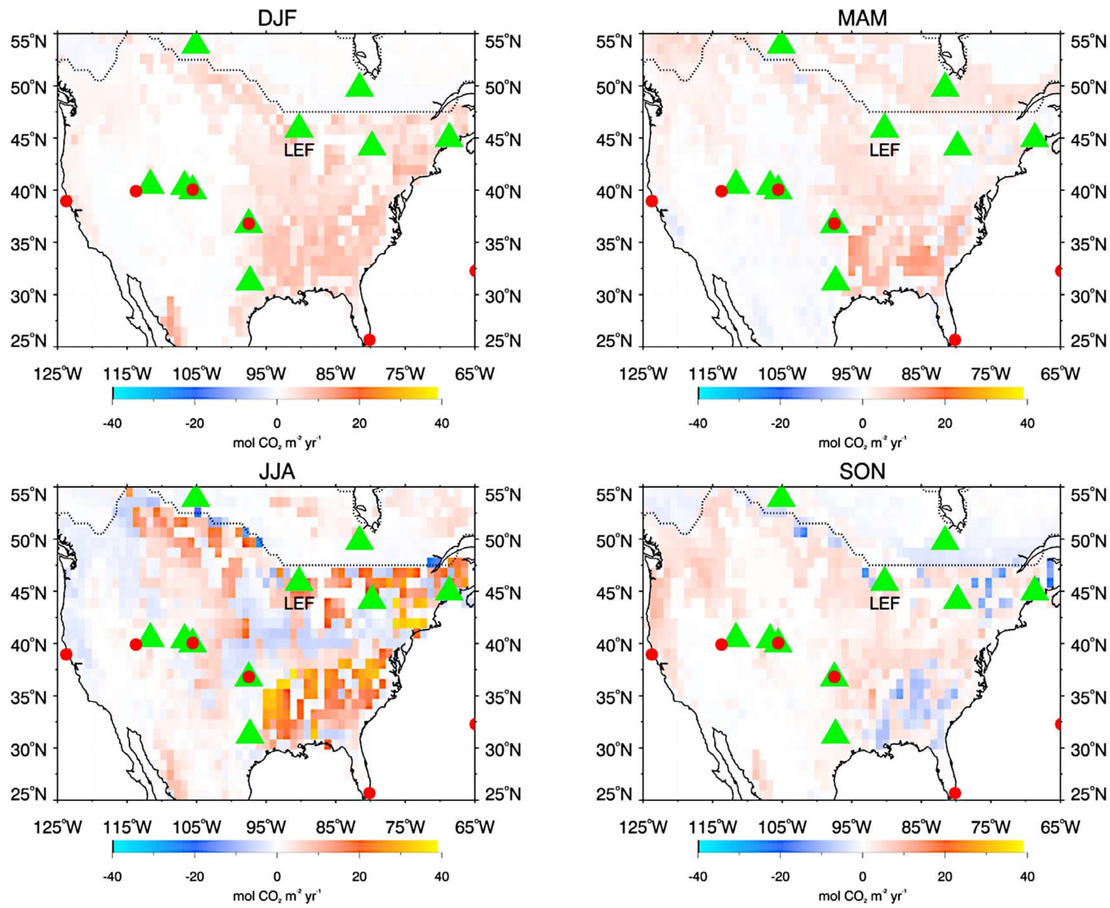

Figure 8. The 1° × 1° seasonal flux differences from the 2004A (+1.5 ppm offset) simulation, test run minus CT2008. Warm colors indicate the test run estimates less surface CO_2 uptake by the terrestrial biosphere than CT2008; cool colors indicate greater surface uptake; and white indicates no difference. Quasi-continuous tower sites (green triangles) and weekly surface measurement sites (red circles) included in CT2008 and the test runs are shown.

is a region with a large NEE and is underobserved, which gives CarbonTracker considerable freedom to adjust fluxes in this region. To test our hypothesis, we repeat the above experiment using the 2004C simulation (+1.5 ppm) test run that excludes the EC observations to the north. The results, which are not shown here, show a slightly larger decrease in surface uptake in the southeastern United States (23 Tg C), while the response in boreal North America is now more than twice as large (44 Tg C).

[23] These experiments suggest that bias in a single measurement record will show up as flux errors elsewhere, especially when those flux regions are not constrained by other sites in the observing network. Future work will include repeating test runs using a more recent CarbonTracker release, which includes quasi-continuous measurements from two tall towers south and east of LEF at West Branch, Iowa (WBI) and Beech Island, South Carolina (SCT) that were not available for CT2008.

3.2. Sensitivity to Biases Exceeding 8 ppm

[24] The 2004B and 2007B simulations were designed to assess the flux response when measurement errors exceed the assigned MDM term [*Peters et al.*, 2007] for the LEF tower observations. In CarbonTracker, the MDM for LEF is set to 3.0 ppm. The inversion system will not assimilate observations that cannot be reproduced by the forecast model to within three times the model-data mismatch error or 9.0 ppm. In practice, measurement biases exceeding ±1 ppm are unlikely to exist in CarbonTracker because of the prerequisite for ongoing comparison experiments to first determine the level of compatibility between independent observations. We find that the response in the estimated flux differences (test minus control runs) is linear until the bias exceeds 3 × MDM error, at which point the flux response becomes nonlinear (Table 3). As the bias increases, more data are rejected by the inversion system. Data rejection is not uniform throughout the year (recall the example of biased observations fitting the model better in summer, for instance) resulting in flux errors that do not scale linearly with the magnitude of the bias or the number of rejected measurements.

[25] In the 2004D simulation, we look at the resulting flux error when all 2004 LEF data are excluded. This result, which is summarized in Table 3 and shown as the horizontal black lines in Figures 4 and 5, shows 40 Tg C less uptake in temperate North America. The intersection between the

Table 3. 2004A, 2004B, and 2004D Simulation Results[a]

Simulation	Introduced Bias (ppm CO_2)	LEF Observations Rejected, 2004 Only	2004 CT Flux Error (Tg C yr^{-1})
2004A	−1.00	2	−80
	+0.25	1	20
	0.50	1	30
	0.75	1	50
	1.00	1	60
	1.50	1	100
	3.00	1	200
	5.00	5	330
2004B	8.50	52	510
	10.00	148	380
	12.00	201	350
	15.00	235	410
2004D	No data		40

[a]For each simulation, the number of 2004 LEF observations rejected during the inversion and the 2004 total annual flux difference (test minus control run) for temperate North America is summarized; 329 afternoon-averaged values are available in 2004.

2004D simulation result and the linear fit to 2004A results for temperate North America (Figure 4) suggests that a measurement bias of ~0.5 ppm results in an artificial flux as large as the flux resulting from the presence of LEF data. This 0.5 ppm value is specific to this study and cannot be generalized since thresholds will depend on network density and proximity to source/sink regions. We can say that there likely exists a bias threshold specific to each measurement record beyond which the observations no longer provide a valid constraint to CarbonTracker. If this is indeed the case, we must continue to strive for the highest level of compatibility between independent records so as to minimize the possibility of approaching these thresholds. A much denser observing network will likely attenuate the sensitivity to bias in any single record unless there exists coherent bias in observations within a region or within a network.

3.3. Sensitivity to Data Gaps

[26] 2004 was selected as the primary test year because of the absence of flux anomalies resulting from a strong El Niño or La Niña event. Unfortunately, there were several measurement gaps in the 2004 LEF record, the largest lasting for approximately one month in June. To evaluate the impact this gap had on this study, we consider 2007, when there were no significant gaps in the LEF measurements. The number of available data sets used in CarbonTracker did not change between 2004 and 2007. Results from the 2007A simulations (Figures 4 and 5, open blue squares) are consistent with those from 2004A, suggesting the gaps in 2004 did not significantly impact our results. The slight difference in slopes is likely due in part to actual differences in net CO_2 exchange between the two years. To remove the year-to-year variability in actual fluxes, we did a single simulation (2007C) where we created gaps in the LEF 2007 data corresponding to the actual measurement gaps existing in the LEF 2004 data. No bias was introduced into the observations. The single open cyan triangle in Figure 4 shows an annual mean reduction in 2007 uptake for temperate North America (~ 40 Tg C) due only to the introduction of gaps. This result is comparable to the flux change when the LEF 2004 data were excluded altogether and is consistent with data reduction scenarios described by *Peters et al.* [2007, 2010]. Gaps in ongoing measurement records are inevitable. They can last days to years and exist in individual records and entire networks. Long-term records sometimes end, creating a “hole” in an observing network. The impact of gaps and intermittent observations on inversion results will depend, in part, on the temporal and spatial density of the observing network.

3.4. Regional Bias and Global Implications

[27] Bias introduced into the LEF record creates inconsistencies with observations made elsewhere. Reduced uptake in North America resulting from the introduced bias at LEF will result in higher predicted CO_2 in Eurasia that is not being observed, thus forcing CarbonTracker to increase uptake in other regions. Figure 9 compares the flux response for North America (temperate plus boreal North America) with extratropical Eurasia (Europe plus temperate and boreal Eurasia). The 2004A results for extratropical Eurasia show an increase in CO_2 surface uptake of ~25 Tg C yr^{-1} for every 1 ppm of bias introduced into the LEF 2004 data. The CarbonTracker response in extratropical Eurasia compensates for ~35% of the response in North America, suggesting CarbonTracker flux estimates in other regions are also affected. This is indeed the case. For example, temperate North Atlantic Ocean and Northern Africa, together and in nearly equal amounts, offset ~20% of the response in North America. The global response is ~20 Tg C yr^{-1} per ppm CO_2 bias and is insignificant relative to the 2004 annual non–fossil fuel global surface flux of ~6500 Tg C yr^{-1} but is consistent with an apparent increase in the NEE flux due to the introduced positive measurement bias.

4. Discussion

[28] Our results suggest that CarbonTracker optimized 2004 annual net surface uptake in temperate North America decreases by ~68 Tg C yr^{-1} (~10% of the estimated North American annual terrestrial uptake) for every 1 ppm of bias introduced into the 2004 LEF CO_2 data. The size of this effect is currently overwhelmed by other sources of uncertainty in CarbonTracker, consistent with the findings of *Rödenbeck et al.* [2006]. For example, the surface flux error caused by a 1 ppm bias at LEF is well within the estimated Gaussian uncertainty of the 2004 annual mean flux for temperate North America (760 ± 390 Tg C yr^{-1}) and estimated flux range (400–1000 Tg C yr^{-1} [*Peters et al.* 2007])

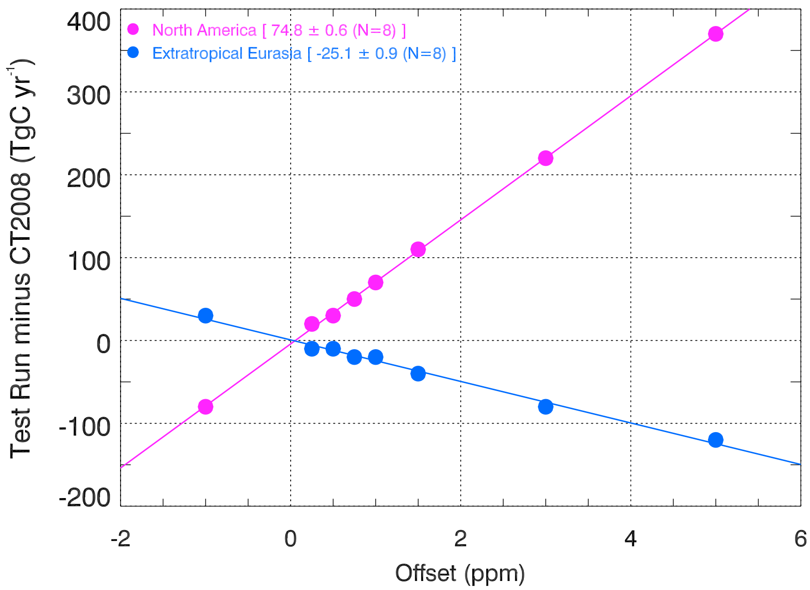

Figure 9. Change in CarbonTracker total annual net CO_2 exchange estimates (test minus control) for North America (temperate plus boreal North America) and extratropical Eurasia (Europe plus temperate and boreal Eurasia) as a function of introduced measurement bias at LEF. The strong anticorrelation between local and hemispheric responses shows that biases propagate beyond the direct footprint of the measurement location because of mass balance constraints on larger scales. Results are from the 2004A simulations.

from CarbonTracker. These estimates reflect uncertainty in both the terrestrial and oceanic prior fluxes, the spatiotemporal data coverage of observations used, and our estimate of TM5′s ability to simulate them. Estimated uncertainty of U.S. CO_2 emissions derived from fossil fuel combustion is ~5% (G. Marland, personal communication) or 80Tg C yr^{-1} based on the 2004 CO_2 emission estimate [*Boden et al.*, 2010].

[29] One might argue that we could relax our measurement compatibility requirements for CarbonTracker since we find that flux errors due to introduced bias are small relative to other sources of uncertainty in CarbonTracker. However, that argument overlooks three facts.

[30] 1. Transport models can be expected to continue to improve and will produce both better reanalysis of historic meteorological fields and mixing processes. The chemical measurements cannot be improved retroactively. Their biases are locked in forever, imposing fixed limitations on future analyses of long-term emissions trends.

[31] 2. Even though models may have significantly different biases when inferring regional fluxes, they tend to agree much better when only the interannual variations of those fluxes are considered [*Bousquet et al.*, 2000]. Our ability to diagnose the response of the regional carbon cycle to climate anomalies (dry or wet summer, warm spring, etc.) would still be significantly impaired when we also have measurement biases that vary over time.

[32] 3. Relevant atmospheric CO_2 signals are also small, and we have to strive for the best signal-to-noise ratio obtainable. As an example, consider a simple carbon mass balance estimate for the United States in which we distribute a net source of 1 Pg C yr^{-1} (Pg C ≡ 10^{15} g carbon) uniformly in space and time in the contiguous 48 states (compare U.S. fossil fuel emissions of ~1.6 Pg C yr^{-1}). The total column-averaged CO_2 mole fraction would increase by only 0.077 ppm per day. If we assume that air transport of the emissions plume over the United States takes 10 days, the increase in the total column would then still only be 0.77 ppm on average downwind of the continent. Only in the unlikely circumstances that the emissions signal were to remain confined to the lowest 3 km over that period, the enhancement would be 2.3 ppm. This illustrates that for robust detection of atmospheric signals, we should continue to strive for the WMO target levels for measurement compatibility of ±0.1 ppm.

[33] The spatial scale over which a bias is applied determines its impact on inferred source/sink distributions. We find a 0.2 ppm bias introduced into the 2004 LEF record results in a ~14 Tg C yr^{-1} error in the CarbonTracker 2004 total annual surface flux estimates for temperate North America. In contrast, *Rödenbeck et al.* [2006] conclude that annual mean biases at remote sites like Mauna Loa and Cape Grim on the order of 0.2 ppm lead to regional flux differences of ~100 Tg C yr^{-1}. While it is difficult to directly compare results from these two studies because of differences in the inversion systems, the spatial and temporal distributions of observations, and the experimental methods, the two studies appear roughly consistent. In this study, a 0.2 ppm constant bias for a single year is introduced into 1 of 13 quasi-continuous records located within North America, among 73 weekly and quasi-continuous sites globally. In *Rödenbeck et al.*, a ~0.2 ppm time-varying bias over 7 years is distributed to either several globally distributed remote marine sites or to entire hemispheric and global networks.

[34] Bias introduced into a single data set from a site in Wisconsin impacts CarbonTracker flux estimates well beyond North America. CarbonTracker flux estimates for Europe and temperate and boreal Eurasia compensate for

the flux change in North America, suggesting the impact of the introduced bias is global. We would expect a similar outcome had we selected a North American site other than LEF. However, the magnitude and extent of the response depends on network density, specifically, how well measurements from any single site are anchored by measurements from other sites. To explore this further, we introduced a +1 and +3 ppm bias into the 2004 Mauna Loa data. Again, we derive a linear flux response to the introduced measurement bias, but because the Mauna Loa Observatory is remote and at an altitude with global-scale representation, the impact on the global total net flux is 3.5 times larger than the LEF (2004A) simulations.

[35] In this paper, we have not yet considered the impact of biases across (1) multiple sites or (2) different observation networks. Such biases do exist and were investigated in more detail by *Rödenbeck et al.* [2006] and by *Peters et al.* [2010] (their case O4). In contrast to these studies, which included more complex and realistic bias structure, we have investigated CarbonTracker's response to a constant bias of varying magnitude at one site. The next steps in this investigation would introduce biases at more than one site and network and would include, for instance, calibration offsets and experimental biases (drifts in reference gases, system leaks, introduction of new detection techniques), similar to those shown in Figures 1 and 2. Our study suggests that even a simple bias cannot be identified using the CarbonTracker inversion system given the current network density and other sources of uncertainty in an inversion. Observational biases can thus only be treated if they are known; this requires diligent long-term monitoring and frequent comparison of atmospheric measurements between labs and instruments.

5. Conclusions

[36] We derive a linear relationship between bias introduced into the Wisconsin tower data and its impact on CarbonTracker surface fluxes for constant biases between −1 and +5 ppm. Results show that the calculated 2004 net surface uptake in temperate North America decreases by 68 Tg C yr^{-1} for every 1 ppm of bias introduced into the LEF 2004 data. The linear response in CarbonTracker flux estimate errors due to observational bias is a feature of the linearity of the inversion system and is independent of the choice of the test site. We have shown that the slope of the sensitivity depends on the region, observation network, and season. Boreal and temperate North America sensitivities increase when we exclude observations in Canada that constrain fluxes near the LEF site. This result demonstrates the value of having multiple sites constraining fluxes from a single region within an observational network.

[37] At present, it is not clear how one would identify measurement and concomitant flux biases in model output. This forces us to conclude that we must continue to employ strategies for identifying bias in observations directly. Indeed, a prerequisite for including CO_2 data into CarbonTracker is a demonstration that observations are compatible to within a few tenths of a ppm CO_2 using ongoing comparison experiments. We will not likely relax measurement compatibility requirements with a much denser cooperative observing network. There will be many more laboratories making observations using different methods and, in some instances, using different calibration scales, or in the case of open-path satellite and FTS measurements, no direct calibration scale. Estimated fluxes derived by inversion systems like CarbonTracker will likely be sensitive to coherent bias between independently operated networks. Future work will focus on assessing the impact of regional network biases on CarbonTracker flux estimates.

[38] **Acknowledgments.** We thank Michel Ramonet for providing the LSCE Mace Head CO_2 observations. The LSCE Mace Head program is part of the Réseau atmosphérique de mesure des composés à effet de serre (RAMCES) network. We are grateful to G. Spain for his invaluable contribution to the NOAA and LSCE sampling programs at Mace Head. We thank our colleagues at CSIRO, NCAR, LBNL, and all the individuals who collect and measure atmospheric samples on which this work is based. Their ongoing conscientious efforts are essential to CarbonTracker. Last, we thank those who have contributed to the development of the components that comprise the CarbonTracker inversion system.

References

Bakwin, P. S., P. P. Tans, C. Zhao, W. Ussler, and E. Quesnell (1995), Measurements of carbon dioxide on a very tall tower, *Tellus, Ser. B*, *47*, 535–549, doi:10.1034/j.1600-0889.47.issue5.2.x.

Bakwin, P. S., K. J. Davis, C. Yi, S. C. Wofsy, J. W. Munger, L. Haszpra, and Z. Barcza (2004), Regional carbon dioxide fluxes from mixing ratio data, *Tellus, Ser. B*, *56*, 301–311, doi:10.1111/j.1600-0889.2004.00111.x.

Boden, T. A., G. Marland, and R. J. Andres (2010), Global, regional, and national fossil-fuel CO_2 emissions, report, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn., doi:10.3334/CDIAC/00001_V2010.

Bousquet, P., P. Peylin, P. Ciais, C. Le Quere, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, *290*(5495), 1342–1346, doi:10.1126/science.290.5495.1342.

Conway, T., P. Tans, L. Waterman, K. Thoning, D. Kitzis, K. Masarie, and N. Zhang (1994), Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network, *J. Geophys. Res.*, *99*(D11), 22,831–22,855, doi:10.1029/94JD01951.

Dee, D., and A. da Silva (1999), Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology, *Mon. Weather Rev.*, *127*, 1822–1834, doi:10.1175/1520-0493(1999)127<1822:MLEOFA>2.0.CO;2.

Engelen, R. J., S. Serrar, and F. Chevallier (2009), Four-dimensional data assimilation of atmospheric CO_2 using AIRS observations, *J. Geophys. Res.*, *114*, D03303, doi:10.1029/2008JD010739.

GLOBALVIEW-CO_2 (2010), *Cooperative Atmospheric Data Integration Project–Carbon Dioxide* [CD-ROM], ESRL, NOAA, Boulder, Colo. [Available at ftp://ftp.cmdl.noaa.gov/ccg/co2/GLOBALVIEW/.]

Huijnen, V., et al. (2010), The global chemistry transport model TM5: Description and evaluation of the tropospheric chemistry version 3.0, *Geosci. Mod. Dev.*, *3*(2), 445–473, doi:10.5194/gmd-3-445-2010.

Jacobson, A. R., N. Gruber, J. L. Sarmiento, M. Gloor, and S. E. Mikaloff Fletcher (2007), A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: I. Methods and global-scale fluxes, *Global Biogeochem. Cycles*, *21*, GB1019, doi:10.1029/2005GB002556.

Krol, M., S. Houweling, B. Bregman, M. van den Broek, A. Segers, P. van Velthoven, W. Peters, F. Dentener, and P. Bergamaschi (2005), The two-way nested global chemistry-transport zoom model TM5: Algorithm and applications, *Atmos. Chem. Phys.*, *5*, 417–432, doi:10.5194/acp-5-417-2005.

Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting (2003), Data and modelling requirements for CO_2 inversions using high-frequency data, *Tellus, Ser. B*, *55*, 512–521, doi:10.1034/j.1600-0889.2003.00029.x.

Machida, T., H. Matsueda, Y. Sawa, Y. Nakagawa, K. Hirotani, N. Kondo, K. Goto, T. Nakazawa, K. Ishikawa, and T. Ogawa (2008), Worldwide measurements of atmospheric CO_2 and other trace gas species using commercial airlines, *J. Atmos. Oceanic Technol.*, *25*, 1744–1754, doi:10.1175/2008JTECHA1082.1.

Masarie, K. A., et al. (2001), NOAA/CSIRO Flask Air Intercomparison Experiment: A strategy for directly assessing consistency among atmospheric measurements made by independent laboratories, *J. Geophys. Res.*, *106*(D17), 20,445–20,464, doi:10.1029/2000JD000023.

Masarie, K., P. Tans, A. Andrews, T. Conway, and A. Crotwell (2009), NOAA comparison activities: Are we closer to the required measurement accuracy?, in *14th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques*, *WMO/TD 1487, GAW Rep. 186*, pp. 33–39, World Meteorol. Org., Geneva.

Miller, C. E., et al. (2007), Precision requirements for space-based X_{CO_2} data, *J. Geophys. Res.*, *112*, D10314, doi:10.1029/2006JD007659.

Morino, I., et al. (2011), Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra, *Atmos. Meas. Tech.*, *4*, 1061–1076, doi:10.5194/amt-4-1061-2011.

Olson, J. S., J. A. Watts, and L. J. Allison (1985), Major world ecosystem complexes ranked by carbon in live vegetation, *Tech. Rep. NDP-017*, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Peters, W., M. Krol, E. J. Dlugokencky, F. Dentener, P. Bergamaschi, G. S. Dutton, P. von Velthoven, J. B. Miller, L. M. P. Bruhwiler, and P. P. Tans (2004), Toward regional-scale modeling using the two-way nested global model TM5: Characterization of transport using SF_6, *J. Geophys. Res.*, *109*, D19314, doi:10.1029/2004JD005020.

Peters, W., et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(48), 18,925–18,930, doi:10.1073/pnas.0708986104.

Peters, W., et al. (2010), Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Global Change Biol.*, *16*, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x.

Popa, M. E., M. Gloor, A. C. Manning, A. Jordan, U. Schultz, F. Haensel, T. Seifert, and M. Heimann (2010), Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, *Atmos. Meas. Tech.*, *3*, 407–427, doi:10.5194/amt-3-407-2010.

Rödenbeck, C., T. J. Conway, and R. L. Langenfelds (2006), The effect of systematic measurement errors on atmospheric CO_2 inversions: A quantitative assessment, *Atmos. Chem. Phys.*, *6*, 149–161, doi:10.5194/acp-6-149-2006.

Tans, P., C. Zhao, and D. Kitzis (2011), The WMO mole fraction scales for CO_2 and other greenhouse gases, and uncertainty of the atmospheric measurements, in *15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurements Techniques*, *WMO TD 1553, GAW Rep. 194*, pp. 101–108, World Meteorol. Org., Geneva.

van der Werf, G.R., J.T. Randerson, L. Giglio, G.J. Collatz, P.S. Kasibhatla, and A. F. Arellano Jr. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, *6*, 3423–3441, doi:10.5194/acp-6-3423-2006.

Vermeulen, A. T., A. Hensen, M. E. Popa, W. C. M. van den Bulk, and P. A. C. Jongejan (2011), Greenhouse gas observations from Cabauw Tall Tower (1992–2010), *Atmos. Meas. Tech.*, *4*, 617–644, doi:10.5194/amt-4-617-2011.

Winderlich, J., H. Chen, C. Gerbig, T. Seifert, O. Kolle, J. V. Lavrič, C. Kaiser, A. Höfer, and M. Heimann (2010), Continuous low-maintenance $CO_2/CH_4/H_2O$ measurements at the Zotino Tall Tower Observatory (ZOTTO) in Central Siberia, *Atmos. Meas. Tech.*, *3*, 1113–1128, doi:10.5194/amt-3-1113-2010.

World Data Centre for Greenhouse Gases (2011), World Data Centre for Greenhouse Gases data summary, *WDCGG 35*, 97 pp, Tokyo.

World Meteorological Organization (2009), *14th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques*, *WMO/TD 1487, GAW Rep. 186*, 145 pp., Geneva.

Wunch, D., et al. (2010), Calibration of the total carbon column observing network using aircraft profile data, *Atmos. Meas. Tech. Discuss.*, *3*, 2603–2632, doi:10.5194/amtd-3-2603-2010.

Zhao, C. L., and P. P. Tans (2006), Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, *J. Geophys. Res.*, *111*, D08S09, doi:10.1029/2005JD006003.

A. Andrews, L. Bruhwiler, T. J. Conway, A. R. Jacobson, K. A. Masarie, J. B. Miller, G. Pétron, and P. P. Tans, Global Monitoring Division, Earth System Research Laboratory, NOAA, 325 Broadway, Boulder, CO 80305, USA. (kenneth.masarie@noaa.gov)

W. Peters, Department of Meteorology and Air Quality, Wageningen University, Droevendaalsesteeg 4, Wageningen NL-6708 PB, Netherlands.

D. E. Worthy, Environment Canada, 4905 Dufferin St., Toronto, Ontario, M3H 5T4, Canada.