Quantifying the observability of CO_2 flux uncertainty in atmospheric CO_2 records using products from NASA's Carbon Monitoring Flux Pilot Project

Lesley Ott[1], Steven Pawson[1], Jim Collatz[2], Watson Gregg[1], Dimitris Menemenlis[3], Holger Brix[4,5], Cecile Rousseaux[1,6], Kevin Bowman[3], Junjie Liu[3], Annmarie Eldering[3], Mike Gunson[3], Stephan R. Kawa[7]

[1]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center
[2]Biospheric Sciences Laboratory, NASA Goddard Space Flight Center
[3]Jet Propulsion Laboratory, California Institute of Technology
[4]University of California, Los Angeles
[5]now at Helmholtz-Zentrum Geesthacht, Institute of Coastal Research, Geesthact, Germany
[6]Universities Space Research Association
[7]Atmospheric Chemistry and Dynamics Laboratory, NASA Goddard Space Flight Center

Submitted to JGR Atmospheres, August, 2014

Abstract

NASA's Carbon Monitoring System (CMS) Flux Pilot Project (FPP) was designed to better understand contemporary carbon fluxes by bringing together state-of-the art models with remote sensing datasets. Here we report on simulations using NASA's Goddard Earth Observing System Model, version 5 (GEOS-5) which was used to evaluate the consistency of two different sets of observationally constrained land and ocean fluxes with atmospheric CO_2 records. Despite the strong data constraint, the average difference in annual terrestrial biosphere flux between the two land (NASA Ames CASA and CASA-GFED) models is 1.7 Pg C for 2009-2010. Ocean models (NOBM and ECCO2-Darwin) differ by 35% in their global estimates of carbon flux with particularly strong disagreement in high latitudes. Based upon combinations of terrestrial and ocean fluxes, GEOS-5 reasonably simulated the seasonal cycle observed at northern hemisphere surface sites and by the Greenhouse gases Observing SATellite (GOSAT) while the model struggled to simulate the seasonal cycle at southern hemisphere surface locations. Though GEOS-5 was able to reasonably reproduce the patterns of XCO_2 observed by GOSAT, it struggled to reproduce these aspects of AIRS observations. Despite large differences between land and ocean flux estimates, resulting differences in atmospheric mixing ratio were small, typically less than 5 ppmv at the surface and 3 ppmv in the XCO_2 column. A statistical analysis based on the variability of observations shows that flux differences of these magnitudes are difficult to distinguish from natural variability, regardless of measurement platform.

1. Introduction

Major weaknesses still exist in our understanding of the processes that control atmospheric carbon dioxide (CO_2) concentrations and as a result, our ability to simulate and predict changes in the Earth system. While the magnitude of the global growth rate of atmospheric CO_2 is well constrained by surface observations (e.g. Conway et al. [1994]), attributing its changes to specific processes remains a challenge given current CO_2 observing capabilities. The Atmospheric Carbon Cycle Inversion Intercomparison (TransCom 3) compared estimates of the carbon budget produced by inverse models using a variety of atmospheric transport fields and CO_2 observations from a global network of surface stations for the period 1992-1996; their results indicated the presence of a larger northern hemisphere (NH) land carbon sink than represented in the a priori flux distribution assumed for the experiment, though the results were strongly influenced by differences in transport fields [Gurney et al., 2002]. The precise location, cause of, and variability of this missing carbon sink remain poorly understood despite the insights provided by intercomparison studies. In addition to uncertainty in the processes governing the contemporary carbon budget, there is evidence that natural land and ocean carbon sinks have decreased over the course of the twentieth century resulting in an increase in the fraction of anthropogenic emissions remaining in the atmosphere, but this is still controversial [Le Quéré et al., 2009; Knorr, 2009].

Model estimates of land and ocean carbon flux are important because they are the only way to understand the underlying processes governing carbon storage and exchange, facilitating an enhanced understanding that may one day contribute to improved abilities to predict changes in the global carbon budget. A number of attempts have been made to compare estimates of carbon flux from different land models. Randerson et al. [2009] compared two biogeochemistry models using a common modeling framework and demonstrated that global carbon sinks differed by a factor of 2 during the 1990s.

Schaefer et al. [2012] and Raczka et al [2013] compared estimates of gross primary production (GPP) and carbon balance among terrestrial ecosystem models over North America and found large differences. Schwalm et al. [2010] compared output from 22 terrestrial biosphere models with data from flux towers in North America and found that models' ability to reproduce observed monthly net ecosystem exchange was poor, though performance was better at forested sites than non-forested sites. Most recently, Huntzinger et al [2012] compared 19 terrestrial biosphere models over North America and found large differences in estimates of GPP (between 12.2 and 32.9 PgC yr^{-1}) and heterotrophic respiration (5.6 to 13.2 PgC yr^{-1}) among models with smaller differences in net ecosystem production (-0.7 to 2.2 PgC yr^{-1}), underscoring the continued uncertainty in land flux processes. Observational estimates of net ecosystem exchange collected worldwide at tower stations provide valuable information to constrain carbon fluxes at local scales (e.g. Lafleur et al. [2003]; Hollinger et al. [2004]; Winderlich et al. [2013]), but are difficult to translate to realistic global estimates because of the spatial heterogeneity of vegetation and limited flux sampling locations.

While ocean carbon fluxes are considered to be reasonably well constrained by the observational database compiled by Takahashi et al. [2002, 2009], certain regions are poorly sampled and models are still needed to understand air-sea exchange processes and predict how they might evolve in the future. As is the case among land models, wide disparity exists among model-derived ocean flux estimates. Doney et al. [2004] showed that errors in model physical processes complicated efforts to compare ocean biogeochemical fields among models while Najjar et al. [2007] documented large circulation differences among 12 global ocean models which influenced export of dissolved organic matter, particularly to the Southern Ocean. Popova et al. [2012] compared five ocean models in the arctic and found that, while primary production was generally consistent among models, the models disagreed over the relative importance of the processes governing production. Uncertainties in the ocean carbon uptake are especially important because they are directly propagated into the inference of the global

terrestrial sink as used by the Global Carbon Project [Le Quere et al, 2013].

It is important to design an Earth observing system that will enable reduction of uncertainties in carbon fluxes and enhance our ability to predict changes in the carbon-climate system. A number of components of the carbon cycle are currently constrained by different types of observations. The Moderate Resolution Imaging Spectroradiometer (MODIS) instruments aboard NASA's Terra and Aqua satellites provide information about vegetation characteristics and fire. Information from the Advanced Very High Resolution Radiometer (AVHRR) instruments aboard NOAA satellites provide a longer term record of vegetation evolution. The Sea-viewing Wide Field-of-view Sensor (SeaWiFS), Suomi National Polar-orbiting Partnership (Suomi NPP), and MODIS both provide information on ocean color and productivity while MODIS and geostationary weather satellites observe the physical ocean state. Weather satellites also enhance our understanding of the carbon cycle by providing information on atmospheric circulation that are necessary for interpreting atmospheric CO_2 observations. Though these remote sensing products inform a number of model-based land and ocean carbon flux estimates, substantial uncertainty remains as documented in the numerous studies cited above.

Atmospheric CO_2 observations provide an important constraint on carbon fluxes, but are limited in space and time. NOAA maintains a network of surface sampling site in remote locations worldwide where trace gas measurements are conducted several times per week as well as a smaller number of observatories and tower sites that provide continuous CO_2 data. Ground-based column CO_2 observations are currently collected at a small number of stations as part of the Total Carbon Column Observing Network (TCCON). Chevallier et al. (2011) demonstrated the utility of column CO_2 data for inferring regional carbon budgets for the first time using data from this sparse network. Several satellite datasets are available from the Atmospheric Infrared Sounder (AIRS) and the Japanese Greenhouses

138 gases Observing SATellite (GOSAT), and, prior to May, 2010, the Scanning Imaging Absorption
139 Spectrometer for Atmospheric Cartography (SCIAMACHY). Satellite CO_2 observations have
140 provided valuable information for constraining surface fluxes, but their impact is limited by their
141 inability to observe in the presence of clouds and aerosols (GOSAT) and low sensitivity to near surface
142 CO_2 mixing ratios (AIRS).
143
144 NASA's Carbon Monitoring System (CMS) Flux Pilot Project (FPP) was designed to better understand
145 the observational constraints placed on carbon fluxes by bringing together state of the art models with a
146 wealth of remote sensing resources. "Bottom up" surface flux estimates were computed by two land
147 and two ocean models for the 2009-2010 period using a consistent set of meteorology input from
148 NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al.
149 [2011]). These results were propagated forward in the atmosphere using NASA's Goddard Earth
150 Observing System Model, version 5 (GEOS-5) constrained by MERRA analyzed meteorological fields.
151 Unlike other model intercomparison projects which span a much larger range of model variables, our
152 intention is not to assess the full range of flux uncertainty and variability; instead, this work brings
153 together a small group of model estimates of flux and atmospheric transport that are tightly constrained
154 by remote sensing datasets and draws upon the expertise of the creators of those datasets to assess flux
155 differences and their manifestation in atmospheric CO_2 records. Here, we report on the results of these
156 simulations with a focus on understanding how uncertainty in observationally constrained land and
157 ocean flux estimates propagates into atmospheric CO_2 and characterizing how differences in flux
158 processes might be observed in the atmosphere by existing platforms. This work is complimentary to
159 the application of "top-down" approaches used to infer the surface fluxes given satellite observations
160 [Liu et al, 2014]. Section 2 of this paper describes the models used in this work while section 3 of this
161 paper describes the comparison with different CO_2 datasets. Section 4 provides a summary and
162 conclusions drawn from this work.

2. Model and Data Background

2.1. Land Biosphere Flux Estimates

2.1.1. CASA-GFED3

The Carnegie-Ames-Stanford-Approach – Global Fire Emissions Database version 3 (CASA-GFED3) derives from Potter et al. [1993], diverging in development since Randerson et al. [1996]. CASA is a light use efficiency type model: net primary production (NPP) is expressed as the product of photosynthetically active solar radiation, a light use efficiency parameter, scalars that capture temperature and moisture limitations, and fractional absorption of solar radiation by the vegetation canopy (FPAR). This latter variable is derived from satellite data.

Fire parameterization was incorporated into the model by van der Werf et al. [2004], producing CASA-GFED, and the model has undergone several revisions (van der Werf et al. [2006, 2010]) leading to its most recent version CASA-GFED3. Input data sets include air temperature, precipitation, incident solar radiation, a soil classification map, and a number of satellite derived products including MODIS vegetation classification, MODIS based burned area, and AVHRR FPAR.

CASA-GFED3 is run at monthly time steps with 0.5° spatial resolution. As part of the CMS FPP, fluxes were computed using MERRA meteorology and FPAR derived from AVHRR NDVI (Tucker et al. [2005]) according to the procedure of Los et al. [2000]. The original 8-km, biweekly AVHRR NDVI was aggregated up to the monthly, 0.5°×0.5°grid by averaging. Model output includes NPP, heterotrophic respiration (Rh), and fire emissions from forest, savanna, deforestation, peat, and agriculture. For this project, fire emissions were disaggregated from monthly to daily using the MODIS active fire products as described in Mu et al. [2011]. Monthly NPP and heterotrophic respiration were disaggregated to 3-hour time intervals following Olsen and Randerson [2004].

2.1.2. NASA Ames CASA

The NASA Ames CASA model (hereafter referred to as Ames CASA) relies on satellite observations of vegetation cover from MODIS as time-series inputs to estimate monthly carbon fluxes from terrestrial ecosystems worldwide. All model algorithms, parameter settings, and land cover data sets used in Ames CASA for CMS flux computations have been documented by Potter et al. [2007].

For the CMS FPP computations of net biosphere fluxes of carbon to the atmosphere, the Ames CASA version documented in Potter et al. [2007] has been modified to use global 0.5^{o} (latitude/longitude) MODIS Enhanced Vegetation Index (EVI) input data (for the years 2000-2010) generated by aggregating monthly 0.05^{o} (~6 km) values. In addition, in cropland areas, 40% of annual NPP carbon is removed each year from the litter decomposition flux pathways and diverted into harvested food products. This is assumed to be re-emitted as a consistent monthly flux (1/12 of the annual cropland harvest carbon total) with a weighted spatial distribution corresponding to the maps of cropland harvest CO_2 emissions developed by Ciais et al. [2007]. Net biosphere fluxes for both CASA-GFED3 and Ames CASA were disaggregated to 3-hour time intervals following Olsen and Randerson [2004]. In contrast to the CASA-GFED3 fluxes, biomass burning is not included in the Ames CASA fluxes computed as part of the CMS FPP. CASA-GFED3 does not include the crop redistribution incorporated in Ames CASA flux estimates.

The Ames CASA fluxes produced as part of the CMS FPP differ from a more recent version of Ames CASA described in Potter et al. [2012] both in the use of different meteorological driver data and in the methodology of process representation.

2.1.3. Differences between Land Flux Estimates

213 Figure 1 shows time series of monthly global flux totals computed by the two land models. Ames
214 CASA estimates a weaker land sink (driven by NH summer) and a stronger source during NH winter
215 compared to CASA-GFED3. These differences result in substantially different annual flux estimates
216 for both years; during 2009 (2010), Ames CASA estimates that the terrestrial biosphere is a 2.5 (2.1)
217 Pg C source while CASA-GFED3 calculates a smaller 0.34 (1.2) Pg C source. The average difference
218 in annual land flux between the two models is 1.7 Pg C. For reference, the uncertainty in land flux
219 between these two models is nearly 20% of the annual mean fossil fuel emissions for the 2009 to 2010
220 period (9 Pg C; Boden et al. [2011]). GFED3 fluxes for 2009 (2010) indicate that fires are a global 2.1
221 (2.9) Pg C source of carbon to the atmosphere while Ames CASA fluxes considered here do not
222 include the effects of fire. In both models, the land biosphere is a source of carbon the atmosphere
223 rather than a sink as indicated by multi-model inversion studies (e.g. Gurney et al. [2002]). Using
224 observed atmospheric CO_2 increases, fossil fuel emission inventories, and ocean models, Le Quéré et
225 al. [2013] estimate the magnitude of the land carbon sink as 3.3 (2.55) Pg C during 2009 (2010). This
226 may be due to the fact that neither model includes several processes thought to be important in
227 explaining the global land sink including CO_2 fertilization (e.g. Bellassen et al. [2011]; Vanuytrecht et
228 al. [2011]; Piao et al. [2013]; Los [2013]), nitrogen deposition (e.g. Esser et al. [2011]; Bala et al.
229 [2013]; Fleischer et al. [2013]; Gerber et al. [2013]), and land use history (e.g. Sentman et al. [2011]).
230 In addition to differences in the magnitude of the land biosphere flux estimates, the Ames CASA and
231 CASA-GFED3 models indicate differences in the phasing of the global seasonal cycle in terrestrial
232 carbon flux to the atmosphere. Ames CASA diagnoses an earlier pattern of drawdown during NH
233 spring and an earlier transition from sink to source during NH autumn in 2009.
234
235 Figures 2 shows the geographic distribution of land fluxes computed by CASA-GFED3 and Ames
236 CASA for four months during 2009. During January, CASA-GFED3 indicates greater drawdown of
237 CO_2 over the southern hemisphere (SH) regions of South America and Africa than does Ames CASA.

238 Fire emissions in equatorial Africa and Australia are also evident in CASA-GFED3 while absent in
239 Ames CASA. In the NH, Ames CASA estimates greater release of CO_2 to the atmosphere along the
240 east coasts of the United States and China. In April, Ames CASA fluxes are generally weak in most
241 locations as the model transitions between winter and summer seasons; areas of western Europe have
242 transitioned from a source in January to a sink of CO_2 in April. In contrast, CASA-GFED3 patterns are
243 generally similar in April and January but fluxes from the land to the atmosphere are slightly larger at
244 most NH locations. During July, CASA-GFED3 estimates greater uptake of CO_2 at most NH locations,
245 particularly at high latitudes, than does Ames CASA. This contrast is also evident over equatorial
246 South America while in mid-latitude regions of the continent, CASA-GFED3 indicates greater release
247 of CO_2 to the atmosphere. The models also differ strongly in their assessment of fluxes over India and
248 Southeast Asia during July with CASA-GFED3 indicating a substantial source and Ames CASA
249 indicating a sink. During October, Ames CASA fluxes are generally much weaker than CASA-GFED3
250 fluxes. Strong fire activity over South America is evident in CASA-GFED3 during October and not
251 included in Ames CASA flux estimates. Differences between the spatial distributions of Ames CASA
252 and CASA-GFED3 fluxes for 2010 (not shown) are qualitatively similar though 2010 was
253 characterized by larger fire emissions in CASA-GFED3.
254
255 **2.2. Ocean Flux Estimates**
256 **2.2.1. NOBM**
257 Global ocean carbon dynamics are simulated by the NASA Ocean Biogeochemical Model (NOBM). It
258 is a three-dimensional representation of coupled circulation, biogeochemical, and radiative processes in
259 the global oceans (Gregg et al. [2003]; Gregg and Casey [2007]). The biogeochemical processes model
260 contains 4 phytoplankton groups, 4 nutrient groups, a single herbivore group, and 3 detrital pools. The
261 phytoplankton groups differ in maximum growth rates, sinking rates, nutrient requirements, and optical
262 properties. Three detrital pools provide for storage of organic material, sinking, and eventual

263 remineralization back to usable nutrients. Radiative transfer calculations provide the underwater
264 irradiance fields necessary to drive growth of the phytoplankton groups, and interact with the heat
265 budget. Carbon cycling involves dissolved organic carbon (DOC) and dissolved inorganic carbon
266 (Gregg et al., 2013). DOC has sources from phytoplankton, herbivores, and carbon detritus, and a sink
267 to DIC. DIC has sources from phytoplankton, herbivores, carbon detritus, and DOC, and
268 communicates with the atmosphere, which can be either a source or sink. The ecosystem sink for DIC
269 is phytoplankton, through photosynthesis. This represents the biological pump portion of the carbon
270 dynamics. The solubility pump portion is represented by the interactions among temperature, alkalinity
271 (parameterized as a function of salinity), silica, and phosphate (parameterized as a function of nitrate).
272 The alkalinity/salinity parameterization utilizes the spatial variability of salinity in the model adjusted
273 to mean alkalinity from the Ocean Model Intercomparison Project (OCMIP; Orr et al. [2001]). The
274 calculations for the solubility pump follow the standards set by the OCMIP.
275
276 NOBM's domain spans from -84^{o} to 72^{o} latitude in increments of 1.25^{o} longitude by $2/3^{o}$ latitude,
277 including only open ocean areas where bottom depth is greater than 200m. The model contains 14
278 vertical layers, in quasi-isopycnal coordinates and is driven by MERRA monthly mean wind stress, sea
279 surface temperature, and shortwave radiation fields. MODIS-Aqua chlorophyll data were assimilated to
280 represent the sum of all phytoplankton components in the model. As part of the CMS FPP, daily
281 NOBM ocean pCO_2 fields were used as input into GEOS-5 and ocean-atmosphere CO_2 fluxes were
282 computed within GEOS-5 using the GEOS-5 atmospheric CO_2 mixing ratios and winds.
283
284 **2.2.2. ECCO2-Darwin**
285 The ECCO2-Darwin Ocean Carbon Cycle Model is based on a global, eddying, ocean and sea ice
286 configuration of the Massachusetts Institute of Technology general circulation model (MITgcm;
287 Marshall et al. [1997a, 1997b]) and on results from two separate projects: the Estimating the

Circulation and Climate of the Ocean, Phase II (ECCO2) Project, which provides a data-constrained estimate of the time-evolving physical ocean state, and the Darwin Project, which provides time-evolving ocean ecosystem variables. Together, ECCO2 and Darwin provide a time-evolving physical and biological environment for carbon biogeochemistry, which is used to compute surface fluxes of carbon at high spatial and temporal resolution.

The ECCO2 model configuration is a cube-sphere grid [Adcroft et al., 2004] with 18-km horizontal grid spacing and 50 vertical levels [Menemenlis et al., 2005a; 2008]. The ECCO2 model configuration includes a dynamic/thermodynamic sea ice model (Losch et al. [2010]; Heimbach et al. [2010]). In a first step, the ECCO2 model configuration was adjusted using a low-dimensional (Green's functions) estimation approach [Menemenlis et al., 2005b]. In a second step, the method of Lagrange multipliers (adjoint method) was used to adjust initial and time-evolving surface boundary conditions [Wunsch and Heimbach, 2007]. Data constraints include sea level anomaly from Jason-1 and OSTM, sea surface temperature from AMSR-E, and temperature and salinity profiles from Argo. This adjoint-based ECCO2 solution is used to drive the Darwin ecosystem model.

The Darwin Project is an initiative to advance the development and application of novel models of marine microbial communities, identifying the relationships of individuals and communities to their environment, connecting cellular-scale processes to global microbial community structure (Follows et al. [2007, 2011]; Dutkiewicz et al. [2009]). The particular configuration used for the CMS FPP includes five phytoplankton functional types (choices based on results from previous versions of the model) and two zooplankton types. The carbon cycle is explicitly included in this configuration, along with those of nitrogen, phosphorus, iron, silica, oxygen, and alkalinity. The carbonate chemistry follows the simplified model proposed by Follows et al. [2006] and air-sea CO_2 exchange is parameterized according to Wanninkhof [1992].

ECCO2-Darwin fluxes produced as part of the CMS FPP are described in more detail in Brix et al. [2014]. GEOS-5 simulations presented here use version 2.1 fluxes. Since the CMS FPP, ECCO2-Darwin fluxes have undergone further development including modification of the piston velocity parameterization.

2.2.3. Differences between Ocean Flux Estimates

NOBM ocean fluxes contain weaker seasonal variations than those produced by the ECCO2-Darwin model (Figure 1). Both models estimate comparable global fluxes for much of the year but during June, July, and August (JJA), ECCO2-Darwin fluxes indicate a weaker ocean sink (or slight source) relative to NOBM; as a result, the 2009 (2010) annual ECCO2-Darwin ocean sink is 2.4 (2.6) Pg C compared to 3.7 (4.0) Pg C from NOBM, a difference of 36% (35%). The Global Carbon Budget [Le Quere et al., 2013] estimates a global ocean flux of 2.57 (2.55) Pg C yr^{-1} for 2009 (2010). NOBM ocean fluxes indicate a stronger ocean sink than the GCB estimate while ECCO2-Darwin estimates differ by less than 10% from GCB in both years.

Figure 3 shows geographic distributions of ocean fluxes computed by NOBM and ECCO2-Darwin for four months in 2009. During January, differences between model estimates of flux are largest in high latitude oceans. NOBM estimates a greater Atlantic Ocean sink north of 30º while ECCO2-Darwin estimates a much stronger ocean sink in the southern ocean than is evident in NOBM. April distributions are generally similar to January, though a decrease in the northern NOBM sink and increase in the ECCO2-Darwin sink relative to January result in spatially inhomogeneous differences north of 30º. During July, ocean flux distributions differ substantially from January; the northern sink in NOBM has weakened and is now evident only in the north Atlantic basin. ECCO2-Darwin's strong southern ocean sink has also weakened while the NOBM southern ocean sink has become stronger.

During October, the NOBM distribution of fluxes is dominated by a strong sink throughout much of the southern hemisphere oceans. ECCO2-Darwin indicates stronger sinks in high latitude regions of both hemispheres, but with less areal extent relative to NOBM. Ocean models also differ in their estimation of coastal fluxes, particularly off the Atlantic coasts of Africa and North America and the Pacific coast of South America.

In regions below 30°S, ECCO2-Darwin estimates an ocean sink 27% (23%) greater than NOBM during 2009 (2010). In addition to differences in magnitude of the sink in this region which includes the Southern Ocean, the seasonal cycle of ocean fluxes is nearly reversed; NOBM estimates peak drawdown in the sink during SH winter months, while ECCO2-Darwin indicates that the sink is at its weakest during these months. In SH Summer, ECCO2-Darwin simulates a maximum in CO_2 drawdown in the region while NOBM simulates a minimum. NOBM, whose domain only extends to 72°N, estimates a sink several times weaker than ECCO2 in NH high latitude oceans. The largest disparities between ocean flux models occur in regions that are rarely observed by ships or are observed only during certain months as shown in Takahashi et al. [2009].

2.3. GEOS-5 Earth System Model

The GEOS-5 Atmospheric General Circulation Model (AGCM) has been developed as a flexible tool to represent the atmosphere on a variety of temporal and spatial scales. It is a central component of the GEOS-5 atmospheric data assimilation system [Rienecker et al., 2008], where it is used with half-degree spatial resolution for meteorological analysis and forecasting [Zhu and Gelaro, 2008] including the production of MERRA which spans the period from 1979 to present [Reinecker et al., 2011]. It has also been developed as a tool for studying atmospheric composition and climate. Ott et al. [2010] used GEOS-5 to examine the impact of Indonesian biomass burning aerosols on atmospheric circulation and Ott et al. [2011] examined the impact of uncertainty in GEOS-5 convection on global carbon monoxide

363 distributions. Model output has also been used by Wang et al. [2009] to compute CO:CO_2 correlations
364 and their impact on inverse modeling.
365
366 The AGCM combines the finite volume dynamical core described in Lin [2004] with the GEOS-5
367 column physics package, summarized in Rienecker et al. [2008]. The model domain extends from the
368 surface to 0.01 hPa and uses 72 hybrid layers that transition from terrain-following near the surface to
369 pure pressure levels above 180 hPa. In this study, the horizontal resolution is 1º x 1.25º (latitude by
370 longitude) and the time step is 30 minutes for physical computations, with more frequent computations
371 of resolved-scale transport in the dynamical core. Trace gases are transported on-line in GEOS-5 using
372 the Lin [2004] dynamical core for resolved scales; turbulent mixing of CO_2 is performed in the same
373 way as for moisture (using the Lock et al. [2000] boundary-layer turbulence module); and using the
374 Relaxed-Arakawa Schubert convective scheme [Moorthi and Suarez, 1992] to represent convective
375 transport. In the present simulations, transport is constrained with MERRA reanalysis fields to ensure
376 consistency with observed meteorology.
377
378 Land biosphere, biomass burning, fossil fuel, and ocean CO_2 fluxes are prescribed in GEOS-5. For the
379 CMS FPP simulations, GEOS-5 was configured to simulate the emission and transport of several
380 different CO_2 tracers representing differing combinations of land and ocean fluxes described in Table 1.
381 In addition to the land and ocean biosphere fluxes provided as part of the CMS FPP, CO_2 emissions
382 from fossil fuels are taken from the Department of Energy's Carbon Dioxide Information Analysis
383 Center (CDIAC; Boden et al. [2011]). Prior to the target 2009-2010 CMS period, CO_2 tracers were
384 spun up from 2000-2008 beginning with a uniform initial condition of 350 ppmv to ensure realistic
385 atmospheric distributions. During the spinup period, land biosphere and biomass burning fluxes from
386 an earlier version of the CASA-GFED model were used in combination with ocean and fossil fuel
387 fluxes from the TransCom Continuous Experiment [Law et al., 2008]. Simulated CO_2 mixing ratios for

December 2008 were calculated at the locations of NOAA Earth Science Research Laboratory (ESRL) marine boundary layer stations [Novelli et al., 1992] and compared with observations; on the basis of this comparison a uniform global offset was subtracted from the simulated CO_2 fields to ensure that global average surface CO_2 concentration was representative of atmospheric conditions at the beginning of the CMS period.

3. Comparison with CO_2 Observational Records

3.1. Comparison with surface CO_2 in situ observations

NOAA's Earth System Research Laboratory (ESRL) Carbon Cycle Greenhouse Gases Group (CCGG) analyzes samples taken weekly at an international cooperative network of surface observing stations (Tans et al., 1990). These data have been used to understand both long term changes in and interannual variability of natural carbon sinks (e.g. Ballantyne et al. [2012]; Conway et al. [1994]). Sites are typically located in remote locations (Figure 4) so that observations represent the background surface CO_2 mixing ratios rather than local source and sink influences.

For comparison to surface stations, the GEOS-5 grid cell containing each station location was sampled at the time that an observation was collected and simulated CO_2 mixing ratios then vertically interpolated to the altitude of the observing station creating model 'pseudo data'. Observations and pseudo-data are averaged over the course of a day (when multiple observations are present) and daily simulated and observed CO_2 mixing ratios at Mauna Loa, Ny Alesund, and Palmer Station (indicated on Figure 4) are shown in Figure 5 for the flux scenarios described in Table 1. One of the primary features evident in this comparison is difference in annual atmospheric growth rate between the flux scenarios. In the beginning of 2009, all scenarios in the GEOS-5 simulations begin from the same CO_2 mixing ratios, but over time differences in the magnitude of the combined land and ocean carbon flux cause the spread in the ensemble of simulations to grow. While this spread in simulations caused by

differing growth rates happens at all stations, it is most evident at Mauna Loa and Palmer Station because of the smaller amplitude of the seasonal cycle at these locations. We calculate growth rates for each flux scenario by first calculating the monthly mean of surface CO_2 mixing ratio at all CCGG marine boundary layer (MBL) sites during 2009 and 2010. For each month, the 2009 MBL monthly mean is subtracted from the corresponding 2010 value and these differences are averaged over all calendar months to estimate the annual mean growth rate shown in Table 1. Using the same method to calculate the annual mean growth rate from observations yields a value of 2.4 ppmv. Flux scenario CG-NO, which combines NOBM and CASA-GFED fluxes, provides the most favorable comparison with the observed growth rate. Both the Ames CASA and ECCO-2 fluxes produce weaker natural sinks than flux scenario CG-NO resulting in greater rates of accumulation of CO_2 in the atmosphere for scenarios AC-NO and CG-ED relative to scenario CG-NO.

In order to examine the ability of the simulations to capture observed spatial and temporal variations in CO_2, we remove the difference in growth rate by subtracting the trends from observations and simulations (Figure 5, right columns). At Mauna Loa, the de-trended time series shows that land biosphere flux differences have a greater impact on simulated mixing ratios than do ocean flux differences. None of the flux scenarios succeeds in reproducing the magnitude of the seasonal cycle observed at Mauna Loa, likely because the NH land sinks in both estimates are too weak or because of deficiencies in model transport. At Ny-Alesund, all model simulations reproduce the seasonal cycle reasonably well though simulations tend to overestimate CO_2, particularly during spring. Differences in seasonal cycle between the Ames CASA and CASA-GFED fluxes are small, but evident at this station in spring and summer months. Ocean flux differences manifest as much smaller mixing ratio differences despite the fact that only one model, ECCO2-Darwin, produces fluxes at this high northern latitude region. At Palmer Station Antarctica, ocean flux differences are larger than at any other station due to the disparity between model estimates of Southern Ocean flux. NOBM overestimates the

438 magnitude of the observed, weak seasonal cycle by 0.5 ppmv while seasonal variations in the ECCO2-
439 Darwin simulation are much too strong resulting in a 5 ppmv overestimate of the seasonal cycle
440 amplitude. The similarity of the CG-NO and AC-NO simulations and large differences between the
441 CG-NO and CG-ED simulations suggest that the seasonal cycle of nearby ocean fluxes more strongly
442 influences the simulation of CO_2 at high latitude SH stations than does the seasonal cycle of land
443 fluxes.
444
445 Figure 6 compares observed and simulated (assuming flux combination CG-NO) monthly mean CO_2
446 mixing ratios at all NOAA ESRL stations collecting substantial amounts of data during the 2009-2010
447 period. Observations at NH stations comprise the bulk of the dataset and show a strong seasonal cycle
448 due to the influence of the land biosphere. The CG-NO simulation also shows a strong seasonal cycle
449 for most stations north of 30°N. At these locations, the model tends to overestimate CO_2, particularly
450 during NH spring and summer, because the CASA-GFED3 land sink is too weak. Comparisons with
451 the Global Carbon Budget presented by Le Quere et al. [2013] indicate an overestimate of global fluxes
452 by the GC-NO flux combination of approximately 3 Pg C which translates into an average, global
453 overestimate in the atmospheric CO_2 of 1.4 ppmv assuming that the excess CO_2 were instantly diffused
454 throughout the mass of the atmosphere. At stations north of 60°N, annual mean CO_2 is, on average,
455 overestimated by 2.3 ppmv during 2010 while stations between the equator and 60°N are overestimated
456 by an average of 1.3 ppmv. CO_2 at northern high latitude stations is overestimated by an average of 4
457 ppmv during spring months and 3 ppmv during summer while NH stations south of 60°N are
458 overestimated by 1.7 ppmv during spring and 1.9 ppmv during summer. The difference in the
459 magnitude of overestimate between NH high and mid-latitude stations may indicate that the CASA-
460 GFED3 land sink underestimate is concentrated in this area or that vertical transport errors are greater
461 at NH high latitudes, resulting in excessive concentration of CO_2 near the surface. In the SH, model
462 errors average 1.7 ppmv during 2010 and show much weaker seasonality.

In Figure 7, we evaluate the ability of GEOS-5 simulations to reproduce the seasonal cycle in CO_2 observed at ESRL stations during 2010. To calculate seasonal cycle errors, monthly means were calculated from observations and model output after de-trending to remove the annual growth rate. We define the amplitude of the seasonal cycle as the difference between the maximum and minimum monthly values. Errors in the months during which minima and maxima in the seasonal cycle occur are also calculated for each station. Means of the amplitude and phasing errors are then calculated over 10 degree latitude bands and only include stations for which data is available in every calendar month during 2010. At NH mid- and high latitude stations, performance is similar for all model flux combinations. Simulations overestimate the magnitude of the seasonal cycle by a few ppmv on average, generally succeed in estimating the summer month during which the minimum should occur, but have more difficulty in estimating the timing of maximum CO_2 ratios in the winter. While errors in the amplitude of the seasonal cycle are small at tropical stations, the model given any flux combination tends to struggle reproducing the observed timing of minima and maxima at SH tropical stations. This is likely due to a combination of the weak seasonal cycle in CO_2 at these locations, errors in the transport of CO_2 from NH locations, and uncertainty in land fluxes in this region that contains dense vegetation. All simulations considered in this study fail to adequately simulate the seasonal cycle at SH high latitude stations. Amplitude errors of several ppmv are often larger than the observed seasonal cycle magnitude. In these locations, ECCO2-Darwin fluxes, which indicate a stronger Southern Ocean sink with greater seasonality relative to NOBM, degrade model performance in terms of both seasonal cycle amplitude and the timing of the maximum in surface CO_2 mixing ratio.

Despite the substantial flux differences among the land and ocean models shown in Figures 1-3, differences in atmospheric CO_2 mixing ratio at remote surface stations tend to be quite subtle as evident in the time series comparisons in Figure 5. To quantitatively determine how atmospheric CO_2

488 observations might be used to discern between these realistic, observationally constrained flux
489 estimates, we follow the method of Huntzinger et al. [2011] who examined the utility of continuous
490 atmospheric flux measurements at a set of tower locations for constraining biospheric flux variability.
491 In order to test the statistical significance of the difference between simulated pseudo-observations, we
492 apply a Chi-square test of variance based on the number of observations per month, the mean squared
493 difference between pairs of pseudo-observations using different flux estimates, and the variance of
494 model-data mismatch [Huntzinger et al., 2011]. Model-data mismatch, an estimate of how closely a
495 model could reproduce an observation, is estimated in this study by estimating the variance of
496 observation residuals around a smooth curve fit to the observations at each site, a method used by
497 Bousquet et al. [1999] and Gurney et al. [2002]. It is important to note that expected model-data
498 mismatch is calculated solely based on variability in the observations. We calculate mean squared
499 difference between simulated CO_2 pairs by first removing the difference in annual growth rate as
500 shown in Figure 5. Qualitatively, this method allows us to assess how well the small atmospheric signal
501 caused by a difference in assumed flux might be distinguished from a background of strong, natural
502 variability.
503
504 Figure 8 shows the difference in monthly mean CO_2 mixing ratio at NOAA ESRL stations resulting
505 from underlying land and ocean flux differences. Differences which are not significantly larger than the
506 model-data mismatch variance at a significance level of 0.05 are indicated by diagonal lines. The
507 largest differences due to land flux are evident at stations in the NH mid- and high latitudes during
508 spring and early summer (2-4 ppmv), when Ames CASA indicates an earlier drawdown of CO_2 by the
509 land biosphere than does CASA-GFED3, and in January through March of 2010 (2-5 ppmv), when the
510 larger land to atmosphere carbon flux in Ames CASA from the latter part of 2009 is most evident. A
511 similar feature is not present in 2009 because it is too close to the initiation of the simulations, but
512 begins to be seen again in December, 2010. Land flux differences are typically less than 2 ppmv at

513 most remote SH locations. At NH locations, land flux differences manifest primarily as a difference in
514 seasonal cycle, consistent with the global flux differences shown in Figure 1. Localized differences,
515 such as those shown in Figure 2 during April and October, are generally not evident in monthly mean
516 CO_2 mixing ratios at NOAA ESRL stations because of their remote locations. The lack of evidence of
517 fire emissions, present in CASA-GFED3 and not Ames CASA, is striking, particularly at SH locations.
518 Analysis of the model's GFED3 biomass burning tracer shows that at all stations except Ascension
519 Island, deviations from the annual mean are smaller than 1 ppmv. At Ascension Island, biomass
520 burning events with greater than 1 ppmv influence only occur on a few days.
521
522 Even in locations and months where substantial differences in CO_2 mixing ratio due to land flux exist,
523 their magnitude is small enough that they are often difficult to distinguish from natural variability
524 evident in the observations. In the NH mid-latitudes, large differences in CO_2 mixing ratio attributable
525 to land flux would not be significantly larger than the expected model-data mismatch because of the
526 large variability evident in the observational records. Differences at NH high latitude stations and in the
527 SH are statistically significant, but only during certain seasons.
528
529 Differences between ocean flux models, which are nearly as large as land flux differences on a global
530 annual basis but exhibit much less seasonality, are even less evident at NOAA ESRL stations. Large
531 differences due to the difference in Southern Ocean sink exist only at stations south of 30°S but are
532 statistically significant in most months. Significant differences between the two ocean models at all
533 stations south of 30°S demonstrate the importance of ocean flux in that region. In this region, NOBM
534 fluxes are more realistic than ECCO2-Darwin fluxes which result in underestimates in surface CO_2
535 mixing ratio and errors in the gradient between tropical and SH high latitude CO_2. Observations
536 suggest a mean difference between Mauna Loa and SH high latitude stations during March, April, and
537 May (MAM) of 6.6 ppmv. While NOBM fluxes reproduce this spatial gradient fairly well, simulating a

538 mean difference of 6.1 ppmv, ECCO2-Darwin fluxes result in a gradient of 8.5 ppmv indicating that
539 Southern Ocean drawdown is too large during certain months. Large differences in ocean CO_2 flux are
540 also present in the NH (Figure 3), but the fact that they occur over smaller regions than in the SH
541 results in smaller mixing ratio differences at NH locations. The combination of smaller mixing ratio
542 differences and greater variability due to biosphere flux and fossil fuel emissions in the NH means that
543 NH ocean flux differences are difficult to distinguish from natural variability. Cold Bay and Shemya,
544 both in Alaska and strongly influenced by marine air, are two exceptions; at these locations differences
545 due to ocean flux are large enough to be distinguished.
546
547 **3.2. Comparisons with HIPPO aircraft observations**
548 Beginning in 2009, the HIAPER (High-performance Instrumented Airborne Platform for
549 Environmental Research) Pole-to-Pole Observations (HIPPO) project has measured atmospheric CO_2,
550 other trace gases, and aerosols in a series of field campaigns. HIPPO missions consist of north-south
551 transects spanning the Pacific ocean from 85°N to 67°S with profiles of atmospheric trace gases every
552 2.2° latitude [Wofsy et al., 2011]. HIPPO data have been used extensively for validation and calibration
553 of remote sensing data products (e.g. Wunch et al. [2010]; Inoue et al. [2013]), for evaluation of
554 atmospheric transport models (e.g. Keppel-Aleks et al. [2013]; Mann et al. [2012]), and to better
555 understand emission and transport processes (e.g. Keppel-Aleks et al. [2012]; Kipling et al. [2013]).
556
557 Three of the five planned HIPPO missions occurred during the 2009-2010 period. Because the first
558 deployment occurred in January, 2009, shortly after the beginning of the CMS target period, our results
559 focus on the second and third HIPPO deployments (HIPPO-2 and HIPPO-3). Both campaigns sampled
560 the Pacific Ocean between 150°E and 90°W. During the second deployment, the majority of samples
561 were collected between 180°W and 150°W while during the third campaign, sampling was focused
562 between 170°W and 140°W. HIPPO data shown are 10 second average data. GEOS-5 pseudo-data are

563 created by sampling the appropriate three-hour model output, interpolating model profiles to the
564 measurement latitude and longitude, and then interpolating resulting profiles vertically to replicate the
565 measurement conditions as well as possible.
566
567 Figure 9 shows a longitudinal cross section comparison between the observed CO_2 data and GEOS-5
568 simulations for HIPPO-2, which occurred during October and November, 2009. Relative to the HIPPO
569 data, GEOS-5 assuming flux combination CG-NO tends to overestimate CO_2 throughout the sampled
570 region of the Pacific. Errors are greater in the NH (1-5 ppmv) than the SH (less than 3 ppmv). The
571 model is able to capture the major features of the observed transect, including elevated near surface
572 CO_2 mixing ratios at NH high latitudes and enhanced CO_2 mixing ratios aloft at 30°S and 30°N.
573 Transects simulated using flux combinations AC-NO and CG-ED are generally similar to the results
574 produced by flux combination CG-NO, but with some notable exceptions. The differences between
575 observations and the CG-NO and AC-NO simulations shows that at NH locations, overestimates in
576 CO_2 are worse when using Ames CASA fluxes compared to CASA-GFED. Near 30°S, the Ames
577 CASA fluxes improve the model's underestimate of near surface CO_2 during SH spring. However,
578 throughout most of the campaign Ames CASA fluxes result in greater errors because the difference in
579 the magnitude of the global land sink is less realistic. The use of ECCO2-Darwin fluxes in flux
580 combination CG-ED results in stronger ocean drawdown at NH high latitudes during autumn. This
581 helps to reduce the overestimate in CO_2 mixing ratios at these locations, but elsewhere, model
582 performance is degraded because of the overall weaker ocean sink.
583
584 During HIPPO-3, conducted during March and April, 2010 (Figure 10), model performance is
585 generally similar to performance during the second deployment. All flux combinations result in
586 overestimates in NH CO_2 but this is worse when using Ames CASA fluxes than when using CASA-
587 GFED fluxes. During March and April, the much stronger southern ocean sink in ECCO2-Darwin is

588 evident in observations near 60°S, resulting in larger errors in this region compared to NOBM.
589
590 We also evaluate the ability of HIPPO-2 and HIPPO-3 data to discern between flux combinations in
591 different regions of the atmosphere following Huntzinger et al. (2011). Model data mismatch is
592 estimated here as the variance of CO_2 observations within 10 degree latitude bins for layers from the
593 surface to 2 km, from 2 to 4 km, from 4 to 6 km, from 6 to 8 km, and from 8 to 9 km. The Chi-squared
594 test statistic is calculated for each of these latitude-altitude regions using the number of observations,
595 the mean squared difference between simulations, and the model data mismatch. Dashed lines on
596 Figures 9 and 10 indicate regions where the test of variance indicates that flux difference induces
597 differences in atmospheric CO_2 mixing ratios that are statistically significant at the 95% confidence
598 level. During HIPPO-2, land flux differences are primarily detectable in the SH mid-latitudes. SH CO_2
599 differences are smaller than those in the NH, but data exhibit much less variability making the
600 differences caused by land flux easier to distinguish from the background variability of measurements.
601 Flux differences in the SH mid-latitudes also tend to be distinguishable through a deep layer of the
602 atmosphere, not only near the surface. Ocean fluxes are difficult to distinguish during the November
603 time frame of HIPPO-2 because the greatest disparity between ocean fluxes is in NH high latitudes
604 where background CO_2 variability is strong. Land fluxes during HIPPO-3, in contrast to HIPPO-2, are
605 detectable only below 4 km and at mid- to high latitudes of both hemispheres. Ocean flux differences
606 are strongly evident in the SH high and mid-latitude locations.
607
608 The HIPPO datasets are also extremely valuable for evaluating the model's ability to simulate vertical
609 gradients in atmospheric CO_2. Figure 11 shows the difference between mean CO_2 near the surface
610 (below 2 km) and mean CO_2 between 6 and 8 km calculated over 5 degree latitude bins for both
611 HIPPO observations and GEOS-5 pseudo-data. During HIPPO-2, GEOS-5 tends to reasonably
612 reproduce the observed interhemispheric gradient both at low levels and aloft. During HIPPO-3, all

613 flux combinations result in the model overestimating the gradient between northern and southern
614 hemisphere CO_2. However, the vertical gradient is reasonably well simulated with the exception of
615 ECCO2-Darwin which results in much stronger negative vertical gradients at mid- and high latitude SH
616 locations than either simulated by NOBM or observed. During HIPPO-3, the earlier spring drawdown
617 in the NH Ames CASA fluxes also results in a slight overestimate in the vertical gradient in CO_2 at
618 most locations. The fact that the interhemispheric gradient in CO_2 is overestimated during HIPPO-3
619 while local vertical gradients appear reasonable could indicate either errors in the assumed source
620 distributions and magnitudes or an error in large scale transport pathways.
621
622 **3.3. Comparison with TCCON column CO_2 observations**
623 The Total Carbon Column Observing Network (TCCON), established in 2004, is a network of ground-
624 based Fourier Transform Spectrometers recording near-infrared direct solar spectra for a number of
625 atmospheric trace gases including CO_2 [Wunch et al., 2011a]. Data are collected continuously during
626 daylight hours when viewing is not obscured by optically thick cloud and aerosol. Observations are
627 estimated to have a precision as high as 0.25% (~ 1 ppmv) under clear sky conditions. In contrast to the
628 NOAA ESRL network, TCCON provides total column CO_2 observations comparable to the
629 observations made by GOSAT. TCCON is also much smaller than the NOAA flask network with 16
630 stations operating during the 2009-2010 study period considered here (Figure 4). TCCON observations
631 are the primary calibration and validation dataset for GOSAT [Wunch et al., 2011b] and their use for
632 flux inference has been demonstrated by Chevallier et al. [2011]. For this analysis, TCCON
633 observations are slightly adjusted to account for laser sampling errors noted by Messerschmidt et al.
634 [2012] and Dohe et al. [2013].
635
636 In order to compare with TCCON observations, GEOS-5 CO_2 fields are sampled at the appropriate
637 observation time and convolved with the averaging kernel and a priori profile information appropriate

638 for each station as described in Wunch et al. (2010). Simulated and observed daily means of column
639 CO_2 are then calculated; the results for four stations are shown in Figure 12. As is expected, column
640 CO_2 observations exhibit smaller seasonal variations than do surface observations. At both high NH
641 mid-latitude stations shown (Garmisch and Lamont), GEOS-5 simulations are able to reasonably
642 reproduce the observed column CO_2 mixing ratios with a slight underestimation in the amplitude of the
643 seasonal cycle (average underestimate over all NH mid- and high latitude stations of 1.4 ppmv for CG-
644 NO fluxes). Differences due to the flux combination assumed are small, typically no more than a few
645 ppmv, but the difference in seasonal cycle between CASA-GFED3 and Ames CASA is evident
646 particularly during NH spring. At the Izana, Tenerife station, GEOS-5 simulations overestimate
647 summer CO_2 mixing ratios relative to observations, a feature also found in the comparison against
648 Mauna Loa surface observations (Figure 5).
649
650 Monthly mean errors between GEOS-5 simulations and TCCON observations (Figure 13) show that
651 model errors are typically less than 3 ppmv when flux combination CG-NO is assumed. The model
652 tends to overestimate CO_2 at NH locations during spring and summer while during winter months,
653 errors are small (less than 1 ppmv at most stations). Model errors at NOAA ESRL NH stations (Figure
654 6) during winter are greater than 5 ppmv at some locations, resulting in small overestimates in seasonal
655 cycle amplitude at the surface (Figure 7). In contrast, the model slightly underestimates the seasonal
656 cycle amplitude at NH TCCON locations. This may be because of errors in simulated transport during
657 winter months resulting in too much CO_2 near the surface, or because of differences in sampling
658 locations between the ESRL and TCCON networks.
659
660 Figure 14 shows differences in column CO_2 mixing ratio at TCCON stations due to the difference in
661 land and ocean fluxes assumed. The seasonality of differences due to differences between Ames CASA
662 and CASA-GFED3 is similar to the analysis of surface observations, but the magnitude of flux

differences manifests as a smaller magnitude atmospheric mixing ratio difference when the column is considered. Despite the smaller total magnitude of the column differences, they are distinguishable from natural variability at most NH stations during late spring and early summer, and winter, 2010. This is largely due to the weaker natural variability in the column data and greater data volume yielded by continuous measurements. The Ames CASA fluxes result in lower CO_2 (0.5-1.5 ppmv) at NH stations during MAM 2009, and higher (2-3 ppmv) CO_2 at NH mid- and high latitude locations during 2010 winter, degrading performance relative to the CASA-GFED3 fluxes.

Detecting ocean flux differences at TCCON locations is considerably more difficult than detecting land flux differences. Data from the Lauder station shows some ability to differentiate between fluxes during some months of 2010, but data are not available at this site during 2009. These results point to the difficulty of detecting ocean CO_2 flux differences with column observations. The nearest surface station, Baring Head (BHD), New Zealand, indicates a higher percentage of data useful for differentiating between ocean fluxes throughout the year. Atmospheric mixing ratio differences caused by ocean flux uncertainty are typically smaller than differences caused by land flux and, as a result, more difficult to separate from natural variability. Additionally, the TCCON network includes no stations further south than Lauder making it impossible to directly observe Southern Ocean flux differences.

3.4. Comparisons with AIRS satellite observations

The Atmospheric Infrared Sounder (AIRS) instrument aboard the Aqua satellite provides a record of mid-tropospheric CO_2 from 2002 to present. Because data are collected during both day and night and retrievals are performed in the presence of clouds, AIRS data provide unprecedented global coverage. Accuracies are reported to be 2 ppmv or better by Chahine et al. [2005] with a nadir footprint of 90 by 90 km^2 [Chahine et al., 2008]. AIRS CO_2 observations have been used to study temporal (e.g. Jiang et

688 al. [2010]; Li et al. [2010]) and spatial (e.g. Ruzmaikin et al. [2012]; Bai et al. [2010]) variability in
689 mid-tropospheric CO_2, to evaluate atmospheric CO_2 simulations (e.g. Feng et al. [2011]), and for data
690 assimilation studies (e.g. Liu et al. [2012]; Engelen et al. [2009]).
691
692 GEOS-5 is sampled according to AIRS version 5 CO_2 observations using level 2 swath data. Simulated
693 CO_2 profiles for the appropriate 3-hour period are interpolated to the location of each observation that
694 passes AIRS quality assurance procedures. AIRS CO_2 weighting functions specific to the observation
695 latitude and background CO_2 mixing ratio are applied as described in Chahine et al. [2008]. Monthly
696 mean simulated and observed CO_2 mixing ratios are then calculated over 5 degree latitude by 5 degree
697 longitude regions.
698
699 Figure 15 presents a comparison between AIRS observations and GEOS-5 simulated CO_2 for October,
700 2009, a month that coincides with the beginning of the HIPPO-2 campaign. There is little agreement
701 between the AIRS observations and the CG-NO GEOS-5 simulation. AIRS indicates peak CO_2 mixing
702 ratios in SH mid-latitudes with minimum values in tropics and sub-tropics of both hemispheres. In
703 contrast, GEOS-5 indicates enhanced mid-tropospheric CO_2 over South America and SH Africa, likely
704 because of the lofting of fire emissions in these regions (Figure 2). GEOS-5 does not indicate the
705 presence of enhanced CO_2 mixing ratios in the mid-latitudes of either hemisphere resulting in a 2-5
706 ppmv underestimate in these regions relative to AIRS. In biomass burning regions, however, GEOS-5
707 overestimates mid-tropospheric CO_2 by 1-3 ppmv relative to AIRS. These results are in sharp contrast
708 to the comparison between HIPPO-2 CO_2 observations and GEOS-5 presented in Figure 9; that
709 comparison indicates that GEOS-5 tends to overestimate mid-tropospheric CO_2 in the 30-60°S band
710 over the Pacific Ocean from 0 to 8 km. The discrepancy between AIRS and GEOS-5 in this region
711 cannot be easily explained by errors in GEOS-5 vertical transport because Figure 11 indicates that
712 GEOS-5 vertical CO_2 gradients throughout the SH differ from observed vertical gradients by less than

713 1 ppmv. NH underestimates of GEOS-5 relative to AIRS also coincide with regions where HIPPO-2
714 data indicate that GEOS-5 is overestimating CO_2 by 2-4 ppmv in the mid-troposphere.
715
716 AIRS data show a significant amount of temporal variability evident in the model data mismatch
717 calculated for each 5 by 5 degree grid box (Figure 15). Model data mismatch is calculated for each grid
718 cell by computing daily means of available data during the 2009-2010 period, fitting a smooth curve
719 through the daily data, and calculating the standard deviation of the residuals, similar to the calculation
720 of model data mismatch for NOAA ESRL and TCCON data. Using this method, expected model data
721 mismatch is typically between 1.5 and 4 ppmv with the largest observed variability in the mid- and
722 high latitudes of both hemispheres. Differences between GEOS-5 simulations assuming different flux
723 combinations are typically small in the mid- and upper troposphere region represented by AIRS
724 observations. The background difference between flux combinations CG-NO and AC-NO, due to the
725 difference in annual growth rate, is 1 ppmv during October. In biomass burning regions in SH Africa
726 and South America, Ames CASA fluxes result in CO_2 mixing ratios 1 ppmv less than CASA-GFED
727 fluxes, or nearly 2 ppmv if the difference in annual growth rate is removed. CASA-GFED results in 2
728 ppmv more CO_2 over equatorial Africa and SE Asia, but if the difference in growth rate is removed,
729 these differences are only 1 ppmv. In tropical regions, where vertical transport of surface fluxes is most
730 rapidly communicated to the mid- and upper troposphere, spatial patterns of atmospheric mixing ratio
731 differences resemble underlying surface flux differences shown in Figure 2. However, the magnitude of
732 these differences (1-2 ppmv) is typically smaller than the expected model data mismatch leading to a
733 lack of ability to discern between flux scenarios with any statistical significance. Ocean flux
734 differences (shown in Figure 15f), which are smaller in magnitude but more consistent throughout the
735 year, are not evident when model simulations are sampled with AIRS pressure weighting functions
736 except as a very slight modification to the north-south CO_2 gradient that is not statistically significant.
737 AIRS is not able to observe the large ocean flux differences in high latitude regions; data are not

738 available south of 60°S and in NH high latitudes, slow vertical transport times produce no spatial
739 signature of the underlying flux differences in the mid- and upper troposphere where AIRS
740 measurements are most sensitive.
741
742 In all months during 2009-2010, AIRS data would have little utility for differentiating between spatial
743 patterns of flux differences. Figure 16 shows the percentage of 5 degree grid boxes per latitude bin that
744 contain statistically significant land and ocean flux differences for each month during the study period.
745 Land flux differences would seldom be detectable and only in a small number of grid cells in the
746 tropics and subtropics. Ocean flux differences are not detectable in any month or location because the
747 magnitude of the mixing ratio differences is always smaller than the variability inherent in the
748 measurements.
749
750 Figure 17 shows a comparison between simulated and observed monthly zonal mean mixing ratios
751 derived from AIRS data and model pseudo-data. GEOS-5 fails to capture the seasonal cycle of zonal
752 mean CO_2 observed by AIRS. GEOS-5 mid-tropospheric CO_2 exhibits a much stronger minimum
753 during the NH growing season (July-October) than does AIRS. AIRS indicates much stronger maxima
754 in the NH high latitudes during spring than indicated by GEOS-5, and a secondary maxima during
755 September that is not simulated by GEOS-5. GEOS-5 tends to produce larger than observed CO_2
756 mixing ratios in the tropics and subtropics throughout the year. In the SH mid-latitudes, GEOS-5
757 underestimates AIRS zonal mean CO_2 with the largest differences in spring and autumn.
758
759 **3.5. Comparisons with GOSAT satellite observations**
760 The Greenhouse gases Observing SATellite (GOSAT), launched in 2009 by the Japanese Aerospace
761 Exploration Agency (JAXA), monitors CO_2 and methane from space by analyzing high resolution
762 spectra of reflected sunlight within several near infrared bands (Kuze et al. [2009]; Yokota et al.,

763 [2009]). The near infrared measurement technique allows for greater sensitivity to CO_2 near the surface
764 than the thermal infrared measurements of AIRS that are sensitive primarily to the mid- and upper
765 troposphere. Retrievals are only performed under cloud-free conditions resulting in a much smaller
766 yield of usable soundings than AIRS. GOSAT observations have been used to examine CO_2 mixing
767 ratios in megacities [Kort et al., 2012] and to estimate regional carbon fluxes (e.g. Takagi et al. [2011];
768 Basu et al. [2013]; Maksyutov et al. [2013]). Several different retrievals of column averaged CO_2 dry
769 air mole fraction (XCO_2) based on GOSAT observations are currently being produced. In this work,
770 we use data produced by NASA's Atmospheric CO_2 Observations from Space (ACOS) effort, a
771 collaboration between the original JAXA GOSAT team and NASA's Orbiting Carbon Observatory
772 (OCO) science team [Crisp et al., 2012]. These data are produced using a modified version of the OCO
773 retrieval algorithm applied to GOSAT observed spectra as described in O'Dell et al. [2012] and have
774 been validated extensively using TCCON measurements (e.g. Wunch et al. [2010, 2011a, 2011b]).
775 Based on retrievals of realistic simulated observations, O'Dell et al. [2012] estimate that observations
776 contain RMS XCO_2 errors of ~1 ppmv and a positive bias of 0.3 ppmv. Dated used here are ACOS
777 version 3.4.
778
779 GEOS-5 is sampled at GOSAT measurement times and locations. Model profiles are interpolated to the
780 twenty atmospheric pressure levels used in the retrieval process and averaging kernels provided as part
781 of the ACOS data product are applied to convolve GEOS-5 mixing ratios with the a priori CO_2 mixing
782 ratio profile. XCO_2 is calculated from this simulated pseudo data profile. Monthly mean simulated and
783 observed CO_2 mixing ratios are then calculated over 5 degree latitude by 5 degree longitude regions as
784 was done for the AIRS data comparison.
785
786 During October, 2009, the spatial distribution of GEOS-5 assuming flux CG-NO compares much more
787 favorably with GOSAT observations (Figure 18) than with AIRS (Figure 15). The distribution of

GOSAT XCO_2 shows fairly uniform mixing ratios globally with slightly larger XCO_2 over South America and Africa. The GEOS-5 distribution is similar though XCO_2 is overestimated by 1-3 ppmv in most locations. This is consistent with HIPPO-2 aircraft observations (Figure 9) which show similar overestimates between 0 and 8 km. Model data mismatch was calculated as described above for AIRS data; observations were binned into 5 degree grid cells, daily means calculated and then the standard deviation of residuals around a smooth curve considered the model data mismatch. Using this technique, estimated model data mismatch is typically between 0.5 and 1 ppmv. However, despite the smaller model data mismatch values, differentiating between land and ocean fluxes remains difficult because of the small number of observations collected. The difference in XCO_2 due to the underlying difference in land flux would be primarily detectable over Africa. Ocean flux differences are too small to be detectable at any location.

A comparison between GOSAT data and GEOS-5 simulations during July 2009 (Figure 19) shows similar features in both models and observations. The model is able to reasonably reproduce the observed spatial distribution of XCO_2 but tends to overestimate XCO_2 by up to 5 ppmv in most mid- and high latitude locations. This is consistent with model overestimates of 3-5 ppmv at surface (Figure 6) and TCCON stations (Figure 13) during NH summer months. Model overestimates are larger during July than during October because the land sink, at its peak during NH summer, is likely weaker in both CASA-GFED and Ames CASA than in reality. The Ames CASA land fluxes, characterized by earlier drawdown of CO_2 in the NH spring, results in smaller CO_2 mixing ratios during July in the NH than CASA-GFED despite the fact that the annual total Ames CASA land sink is smaller. Though the largest differences in land flux during July are in the NH high latitudes (Figure 3), the resultant atmospheric mixing ratio difference is not generally detectable at these locations. Instead, land flux differences are most readily observed over NH ocean locations and over North Africa because the observations exhibit much less variability in these locations. As in October, ocean flux differences are

not large enough to be detectable at any locations.

Figure 16 also shows the percentage of 5 degree grid boxes per latitude bin that contain statistically significant ($p<0.05$) land and ocean flux differences for each month when GOSAT data are considered. In contrast to AIRS and in spite of very sparse sampling, GOSAT observations would be able to distinguish spatial pattern differences in atmospheric CO_2 due to differing land fluxes in both hemispheres primarily during their spring and summer months. Ocean flux differences are statistically significant in only a small number of grid cells in the SH tropics.

Simulated zonal mean XCO_2 from GEOS-5 agrees much more favorably with GOSAT data (Figure 20) than with AIRS (Figure 17) as was the case when spatial distributions of CO_2 during October 2009 were considered. GEOS-5 tends to overestimate XCO_2 throughout the year with the largest degree of overestimation during NH spring and summer. The use of Ames CASA instead of CASA GFED land fluxes help to improve the overestimation during spring, but degrades the comparison with observations during NH winter. Atmospheric CO_2 differences caused by differing land fluxes are detectable when zonal means are considered primarily during spring and summer months in both hemispheres. Ocean flux differences are too small to be detected by GOSAT observations in all months and in all locations.

4. Summary and Conclusions

As part of NASA's CMS Flux Pilot Project, the GEOS-5 GCM was used to simulate atmospheric CO_2 mixing ratios during 2009 and 2010 using two sets of model-based land and ocean flux estimates. All land and ocean flux estimates were informed by multiple satellite datasets and compared with numerous observations with the goal of better understanding the constraint on carbon flux provided by current observing systems. Despite the strong data constraint, the two land models used in this work

838 differed by an average of 1.7 Pg C in their estimate of the global land carbon sink during 2009 and
839 2010. Both land flux models estimate the land biosphere as a net source of carbon to the atmosphere
840 rather than a ~ 3 Pg C sink as indicated by the Global Carbon Budget [Le Quere et al., 2013]. While the
841 presence of a missing land sink was indicated by TRANSCOM models over 10 years ago (e.g. Gurney
842 et al. [2002]) and much progress has since been made in understanding the processes which may
843 account for the disparity between forward and inverse model flux estimates, the CMS FPP results show
844 that estimating realistic net land carbon fluxes at both global and regional scales remains a challenge
845 for models, even those informed by remote sensing observations. Ocean models differ by 1.4 Pg C per
846 year in global carbon flux. In ocean waters south of 30°S, model flux differences are 25% as large as
847 the mean regional flux and the phasing of the seasonal cycle also differs substantially. The magnitude
848 of these differences underscores the continued uncertainty surrounding the ability of natural carbon
849 reservoirs to compensate for increasing fossil fuel emissions.
850
851 Despite considerable differences in global flux, the performance of GEOS-5 simulations is generally
852 similar because of the strong role played by meteorological transport. GEOS-5 simulations assuming
853 all flux combinations tended to overestimate surface CO_2 mixing ratios in the NH, particularly during
854 spring and summer seasons when errors are 3-5 ppmv. The amplitude and phasing of the seasonal cycle
855 at most NH locations is reasonably well simulated though the model tends to slightly overestimate the
856 amplitude of the seasonal cycle at the surface. All flux combinations examined in this study struggle to
857 adequately reproduce the timing of the observed seasonal cycle at SH stations and at mid- and high
858 latitude SH locations and strongly overestimate the magnitude of seasonal variability by as much as 5
859 ppmv in some locations. Because the seasonal cycle is much smaller at SH mid- and high latitude
860 compared to NH locations, errors of this magnitude are greater than the observed seasonal cycle
861 amplitude.
862

863 Comparisons with HIPPO observations provide the opportunity to evaluate the realism of simulated
864 vertical mixing processes. As expected from flux and surface CO_2 comparisons, GEOS-5 tends to
865 overestimate CO_2 mixing ratios relative to aircraft observations, particularly in the NH mid- and high
866 latitudes. Differences between mean near surface CO_2 and CO_2 observed between 6 and 8 km show the
867 model succeeds in simulating realistic vertical gradients in CO_2 during October-November, 2009 and
868 March-April, 2010. The HIPPO data comparisons indicate that during these months and over the
869 Pacific, model errors in surface and column CO_2 are not likely to be attributable to vertical transport
870 errors.
871
872 GEOS-5 simulations succeed in capturing the major features of column CO_2 observations from
873 TCCON and GOSAT though the model tends to overestimate CO_2 by 3-4 ppmv during NH summer
874 regardless of the combination of fluxes used. GEOS-5 is also able to reproduce the spatial patterns of
875 XCO_2 observed by GOSAT. When compared with AIRS mid-tropospheric CO_2 mixing ratios,
876 however, the model fails to represent either observed spatial distributions or seasonal cycle information.
877
878 Comparisons between GEOS-5 and surface, aircraft, TCCON, and GOSAT data reveal a consistent
879 picture of the capabilities of contemporary models to reproduce atmospheric CO_2 observations. Models
880 tend to overestimate CO_2 because the combined land and ocean flux assumed is too weak.
881 Overestimates at the surface are largest during the NH summer months, consistent with an
882 underestimate in the strength of land biosphere sink that is most active in this region and during this
883 season. HIPPO observations show that vertical gradients are reasonably well simulated, meaning that
884 excess CO_2 in the model atmosphere is distributed through the deep layer (typically between the
885 surface and 8 km) sampled by the aircraft. Total column CO_2 measurements (TCCON and GOSAT)
886 are overestimated by GEOS-5 in the same locations and seasons as indicated by comparisons with in
887 situ measurements, but by a slightly smaller amount. It is more difficult to reconcile AIRS mid-

888 tropospheric CO_2 observations with other observational records and GEOS-5. While AIRS vertical
889 weighting functions peak in the mid-troposphere, they are strongly sensitive to the upper
890 troposphere/lower stratosphere (UTLS). Depending on the latitude of the observation, 60-71% of the
891 observed signal comes from pressures lower than 500 hPa. In order for GEOS-5 simulations to both
892 agree with TCCON and GOSAT and disagree with AIRS CO_2 observations, the model would have to
893 have large errors in the UTLS including overestimates in the tropics and alternating under- and
894 overestimates in the NH high latitudes. While there is no evidence of this type of model bias in the
895 comparisons with HIPPO presented here, sampling above 8 km is fairly sparse. More work is needed to
896 thoroughly evaluate the consistency of AIRS and other CO_2 data products, especially if they are to be
897 used together in joint assimilation and inversion frameworks.
898
899 Despite differences between land and ocean flux estimates greater than 1 Pg C, resulting differences in
900 atmospheric mixing ratio at remote surface sites are small, typically less than 5 ppmv at the surface and
901 3 ppmv in the column, and difficult to distinguish from natural observed variability. At NOAA ESRL
902 surface stations, the difference between the Ames CASA and CASA-GFED land fluxes manifests most
903 clearly in the atmosphere during NH spring and winter months. Outside of NH high latitude locations,
904 CO_2 mixing ratio differences are rarely statistically significant when the natural variability of
905 observations is considered. TCCON observations succeed in detecting seasonal differences between
906 land flux estimates in NH winter and spring months. Though flux differences result in smaller mixing
907 ratio differences in the column, the continuous monitoring strategy of TCCON produces larger
908 quantities of data, which facilitates separation of flux differences from natural variability. AIRS
909 satellite observations are unable to discern between land flux models because the variability of
910 observations is larger than the magnitude of the resulting simulated mixing ratio differences in the mid-
911 and upper troposphere. GOSAT observations are better suited to observing flux differences because of
912 greater sensitivity near the surface and they show an ability to distinguish between land flux estimates

913 during summer months. However, GOSAT's ability to differentiate between land flux estimates
914 remains limited by sparse coverage.
915
916 Ocean flux differences, characterized by large disparities at high latitudes of both hemispheres, are
917 primarily distinguishable from natural variability at SH mid- and high latitude surface stations. HIPPO
918 aircraft observations also succeed in differentiating Southern Ocean flux differences from natural
919 variability. While in situ observations show an ability to discern between ocean flux estimates, remote
920 sensing techniques fail for several reasons. The sparse TCCON network contains no stations below
921 45°S, the region where flux uncertainty is greatest and where surface observations show the greatest
922 ability to differentiate between ocean flux estimates. As was the case with land flux, ocean flux
923 estimates produce mixing ratio differences that are smaller than the temporal variability inherent in
924 AIRS data. GOSAT observations, which have smaller single observation errors than AIRS and greater
925 sensitivity near surface, do not observe ocean regions south of 40°S. The difference between the ocean
926 flux models included in the CMS FPP highlights the larger uncertainty in how high latitude ocean
927 carbon storage has changed in recent decades and how it may continue to change in response to future
928 climate change. The inability of current remote sensing observations to detect the large differences in
929 ocean flux presented here highlights the need for continued in situ observations and the development of
930 remote sensing techniques which have the potential to increase data yield over high latitude regions.
931
932 Small mixing ratio differences resulting from flux differences make it difficult to assess the
933 performance of the individual flux estimates. While CASA-GFED compares more favorably than
934 Ames CASA to the total land flux estimate from the Global Carbon Budget [Le Quere et al., 2013], it
935 still diagnoses the land biosphere as a net source of CO_2 to the atmosphere rather than a sink. NOBM
936 overestimates the magnitude of the ocean sink relative to the Global Carbon Budget and ECCO2-
937 Darwin, but because both land flux estimates underestimate the magnitude of the global sink, the CG-

938 NO simulation produces the most realistic atmospheric growth rate when compared with NOAA
939 surface observations. The main land flux difference that manifests in atmospheric CO_2 mixing ratios is
940 the difference in seasonal cycle between CASA-GFED and Ames CASA. While the earlier spring
941 drawdown in Ames CASA results in slightly better comparisons with surface observations during the
942 spring, earlier release of CO_2 from the land to the atmosphere in autumn degrades comparisons during
943 winter. Small-scale differences between CASA-GFED and Ames CASA, including the presence of fire
944 emissions (CASA-GFED) or redistribution of crop CO_2 (Ames CASA), are difficult to evaluate
945 because these differences are not readily evident in the available observational CO_2 records. While
946 ECCO2-Darwin produces a global ocean sink magnitude nearly equal to the Global Carbon Budget
947 estimate, comparisons with surface and aircraft data show that it is less realistic than NOBM over the
948 Southern Ocean where it diagnoses a larger sink with stronger seasonal variations.
949
950 While all flux combinations perform reasonably well in reproducing observed seasonal cycles and
951 spatial gradients in CO_2, the troubling implication of this agreement between simulations is that even
952 differences between flux models on the order of Pg C are difficult to disentangle using current
953 atmospheric CO_2 observations. Ocean flux differences are particularly difficult to discern because they
954 are smaller and tend to occur in high latitude regions that are poorly observed by current remote
955 sensing platforms. The use of aircraft campaigns, such as HIPPO, provide a valuable complement to
956 existing long term carbon monitoring strategies and may become more important in the future as
957 natural carbon reservoirs in remote high latitude locations respond to a changing climate.
958
959 Observations from OCO-2 are expected to greatly improve data yield over much of the globe, which
960 would allow for a greater ability to use space-based observations for flux discrimination. However,
961 OCO-2 will also be limited in its ability to observe high latitude locations meaning that it will not be
962 able to directly observe many locations associated with large flux uncertainty throughout the year. The

963 ASCENDS mission, which focuses on active rather than passive observations of CO_2, would improve
964 the ability of GOSAT and OCO-2 to observe high latitude locations, but is not likely to launch until the
965 2020s. These observational challenges underscore the need for a variety of types of CO_2 observations
966 to help fill the gaps left by satellite observations and to provide additional, complementary information
967 to maximize the impact of satellite observations in regions where they are available.
968
969 The small differences in atmospheric CO_2 mixing ratio due to flux also underscore the importance of
970 quality meteorological analyses and models. The ability to reasonably simulate small gradients in
971 atmospheric CO_2 mixing ratio and to successfully track the transport of air parcels from the surface to
972 the locations and altitudes at which measurements occur is key to the success of atmospheric inversion
973 studies which seek to reduce uncertainty in natural carbon sinks using atmospheric CO_2 observations.
974 Work is ongoing as part of the CMS project to better understand and quantify transport model errors.
975

976 **Acknowledgements**

977 This work was supported by funding from NASA's CMS FPP and Phase One activities. MERRA data
978 used in during the CMS FPP have been have been provided by the GMAO at NASA Goddard Space
979 Flight Center through the NASA GES DISC online archive. Surface observations were provided by
980 NOAA ESRL. TCCON data were obtained from the TCCON Data Archive, operated by the California
981 Institute of Technology from the website at http://tccon.ipac.caltech.edu/. HIPPO data were retrieved
982 from the HIPPO data archive at http://hippo.ornl.gov/dataaccess. AIRS is managed by NASA's Jet
983 Propulsion Laboratory with data provided through the NASA GES DISC. ACOS data were provided
984 by Chris O'Dell from Colorado State University.
985

References

Adcroft, A., J. Campin, C. Hill, and J. Marshall (2004), Implementation of an atmosphere-ocean general circulation model on the expanded spherical cube, *Monthly Weather Review, 132*, 2845–2863.

Bai W., X. Zhang, and P. Zhang (2010), Temporal and spatial distribution of tropospheric CO2 over China based on satellite observations, *Chinese Science Bulletin, 55*(31), 3612-3618.

Bala, G., N. Devaraju, R. K. Chaturvedi, K. Caldeira, and R. Nemani (2013), Nitrogen deposition: how important is it for global terrestrial carbon uptake?, *Biogeosciences, 10*, 7147-7160, doi:10.5194/bg-10-7147-2013.

Ballantyne, A. P., C. B. Alden, J. B. Miller, P. P. Tans, and J. W. C. White (2012), Increase in observed net carbon dioxide uptake by land and oceans during the last 50 years, *Nature, 488*, 70-72.

Basu, S., S. Guerlet, A. Butz, S. Houweling, O. Hasekamp, I. Aben, P. Krummel, P. Steele, R. Langenfelds, M. Torn, S. Biraud, B. Stephens, A. Andrews, and D. Worthy (2013), Global CO2 fluxes estimated from GOSAT retrievals of total column CO2, *Atmospheric Chemistry and Physics Discussions, 13*, 4535-4600, doi:10.5194/acpd-13-4535-2013.

Bellassen, V., N. Viovy, S. Luyssaert, G. Le Maire, M.-J. Schelhaas, and P. Ciais (2011), Reconstruction and attribution of the carbon sink of European forests between 1950 and 2000, *Global Change Biology, 17*, 3274–3292. doi: 10.1111/j.1365-2486.2011.02476.x.

Boden, T. A., G. Marland, and R. J. Andres (2011), Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2011.

Bousquet, P., P. Ciais, P. Peylin, M. Ramonet, and P. Monfray (1999), Inverse modeling of annual atmospheric CO2 sources and sinks: 1. Method and control inversion, *Journal of Geophysical Research*, *104*(D21), 26161–26178, doi:10.1029/1999JD900342.

Brix, H., D. Menemenlis, C. Hill, S. Dutkiewicz, O. Jahn, D. Wang, K. Bowman, and H. Zhang (2014), Using Green's Functions to initialize and adjust a global, eddying ocean biogeochemistry general circulation model, *Ocean Modelling*, submitted.

Chahine, M., C. Barnet, E. T. Olsen, L. Chen, and E. Maddy (2005), On the determination of atmospheric minor gases by the method of vanishing partial derivatives with application to CO2, *Geophysical Research Letters, 32*, L22803, doi:10.1029/2005GL024165.

Chahine, M. T., L. Chen, P. Dimotakis, X. Jiang, Q. Li, E. T. Olsen, T. Pagano, J. Randerson, and Y. L. Yung (2008), Satellite remote sounding of mid-tropospheric CO2, *Geophysical Research Letters, 35*, L17807, doi:10.1029/2008GL035022.

Chevallier, F., et al. (2011), Global CO_2fluxes inferred from surface air-sample measurements and from TCCON retrievals of the CO_2 total column, *Geophysical Research Letters, 38*, L24810, doi:10.1029/2011GL049899.

Ciais, P., P. Bousquet, A. Freibauer, and T. Naegler (2007), Horizontal displacement of carbon associated with agriculture and its impacts on atmospheric CO_2, *Global Biogeochemical Cycles,* doi:10.1029/2006GB002741.

Conway, J., P. P. Tans, L. S. Waterman, K. W. Thoning, D. R. Kitzis, K. A. Masarie, and N. Zhang, (1994), Evidence of interannual variability of the carbon cycle from the NOAA/CMDL global air sampling network, *Journal of Geophysical Research*, *99*, 22831-22855.

Crisp, D., et al. (2012), The ACOS CO2 retrieval algorithm – Part II: Global XCO2 data characterization, *Atmos. Meas. Tech., 5*, 687-707, doi:10.5194/amt-5-687-2012.

Doney, S. C., et al. (2004), Evaluating global ocean carbon models: The importance of realistic physics, *Global Biogeochemical Cycles, 18*, GB3017, doi:10.1029/2003GB002150.

Dutkiewicz, S., M. Follows, and J. Bragg (2009), Modeling the coupling of ocean ecology and biogeochemistry, *Global Biogeochem.Cycles, 23*, GB4017.

Engelen, R. J., S. Serrar, and F. Chevallier (2009), Four-dimensional data assimilation of atmospheric CO2 using AIRS observations, *Journal of Geophysical Research, 114*, D03303, doi:10.1029/2008JD010739.

Esser, G., J. Kattge, and A. Sakalli (2011), Feedback of carbon and nitrogen cycles enhances carbon sequestration in the terrestrial biosphere, *Global Change Biology, 17*, 819–842, doi: 10.1111/j.1365-2486.2010.02261.x.

Feng, L., P. I. Palmer, Y. Yang, R. M. Yantosca, S. R. Kawa, J.-D. Paris, H. Matsueda, and T. Machida (2011), Evaluating a 3-D transport model of atmospheric CO2 using ground-based, aircraft, and space-borne data, *Atmospheric Chemistry and Physics*, 11, 2789-2803, doi:10.5194/acp-11-2789-2011

Fleischer, K., K. T. Rebel, M. K. van der Molen, J. W. Erisman, M. J. Wassen, E. E. van Loon, L. Montagnani, C. M. Gough, M. Herbst, I. A. Janssens, D. Gianelle, and A. J. Dolman (2013), The contribution of nitrogen deposition to the photosynthetic capacity of forests, *Global Biogeochemical Cycles*, *27*, 187–199, doi:10.1002/gbc.20026.

Follows, M., T. Ito, and S. Dutkiewicz (2006), On the solution of the carbonate chemistry system in ocean biogeochemistry models, *Ocean Modelling, 12*, 290–301.

Follows, M., S. Dutkiewicz, S. Grant, and S. Chisholm (2007), Emergent biogeography of microbial communities in a model ocean, *Science, 315*, 1843–1846.

Follows, M. and S. Dutkiewicz (2011), Modeling diverse communities of marine microbes, *Annu. Rev. Marine Science, 3*, 427–451.

Gerber, S., L. O. Hedin, S. G. Keel, S. W. Pacala, and E. Shevliakova (2013), Land use change and nitrogen feedbacks constrain the trajectory of the land carbon sink, *Geophysical Research Letters, 40,* 5218–5222, doi:10.1002/grl.50957.

1083 Gregg, W. W., P. Ginoux, P. S. Schopf, and N. W. Casey (2003), Phytoplankton and Iron: Validation
1084 of a global three-dimensional ocean biogeochemical model, *Deep-Sea Research II, 50*, 3143-3169.
1085
1086 Gregg, W. W. and N. W. Casey (2007), Modeling coccolithophores in the global oceans, *Deep-Sea*
1087 *Research II, 54*, 447-477.
1088
1089 Gregg, W. W., N. W. Casey, and C. S. Rousseaux (2013), Global surface ocean estimates in a model
1090 forced by MERRA, NASA Global Modeling and Assimilation Series, M. Suarez, ed., NASA
1091 Technical Memorandum 2012-104606, Vol. 31, 32 pp.

1092 Gurney, K. R., et al. (2002), Towards robust regional estimates of CO2 sources and sinks using
1093 atmospheric transport models, *Nature, 415*, 626-630.
1094
1095 Heimbach, P., D. Menemenlis, M. Losch, J. Campin, and C. Hill (2010), On the formulation of sea-ice
1096 models. Part 2: Lessons from multi-year adjoint sea ice export sensitivities through the Canadian Arctic
1097 Archipelago, *Ocean Modelling, 33*, 145–158.
1098
1099 Hollinger, D. Y., et al. (2004), Spatial and temporal variability in forest–atmosphere CO_2 exchange,
1100 *Global Change Biology*, 10: 1689–1706. doi: 10.1111/j.1365-2486.2004.00847.x
1101
1102 Huntzinger, D. N., S. M. Gourdji, K. L. Mueller, and A. M. Michalak (2011), The utility of continuous
1103 atmospheric measurements for identifying biospheric CO2 flux variability, *Journal of Geophysical*
1104 *Research, 116*, D06110, doi:10.1029/2010JD015048.
1105
1106 Huntzinger, D. N., et al. (2012), North American Carbon Program (NACP) regional interim synthesis:
1107 Terrestrial biospheric model intercomparison, *Ecological Modelling*, *232*, 144-157.
1108
1109 Inoue, M., et al. (2013), Validation of XCO2 derived from SWIR spectra of GOSAT TANSO-FTS with
1110 aircraft measurement data, *Atmospheric Chemistry and Physics Discussions, 13*, 3203-3246,
1111 doi:10.5194/acpd-13-3203-2013.
1112
1113 Jiang, X., M. T. Chahine, E. T. Olsen, L. L. Chen, and Y. L. Yung (2010), Interannual variability of
1114 midtropospheric CO2 from Atmospheric Infrared Sounder, *Geophysical Research Letters, 37*, L13801,
1115 doi:10.1029/2010GL042823.
1116
1117 Keppel-Aleks, G., et al. (2013), Atmospheric carbon dioxide variability in the Community Earth
1118 System Model: Evaluation and transient dynamic during the 20th and 21st centuries, *Journal of*
1119 *Climate, 26*, 4447-4475, doi:10.1175/JCLI-D-12-00589.1.

1120 Keppel-Aleks, G., (2012), The imprint of surface fluxes and transport on variations in total column
1121 carbon dioxide, *Biogeosciences, 9*, 875-891, doi:10.5194/bg-9-875-2012.
1122
1123 Kipling, Z., P. Stier, J. P. Schwarz, A. E. Perring, J. R. Spackman, G. W. Mann, C. E. Johnson, C. E.,
1124 and P. J. Telford (2013), Constraints on aerosol processes in climate models from vertically-resolved
1125 aircraft observations of black carbon, *Atmospheric Chemistry and Physics*, *13*, 5969-5986,
1126 doi:10.5194/acp-13-5969-2013.
1127
1128 Kort, A. E., C. Frankenberg, C. E. Miller, and T. Oda (2012), Space-based observations of megacity
1129 carbon dioxide, *Geophysical Research Letters, 39*, L17806, doi:10.1029/2012GL052738.

Kuze, A., S. Hiroshi, N. Masakatsu, and H. Takashi (2009), Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, *Applied Optics*, *48*, 6716-6733, doi:10.1364/AO.48.006716.

Lafleur, P. M., N. T. Roulet, J. L. Bubier, S. Frolking, and T. R. Moore (2003), Interannual variability in the peatland-atmosphere carbon dioxide exchange at an ombrotrophic bog, *Global Biogeochem. Cycles*, 17, 1036, doi:10.1029/2002GB001983, 2.

Law, R. M., et al. (2008), TransCom model simulations of hourly atmospheric CO2: Experimental overview and diurnal cycle results for 2002, *Global Biogeochem. Cycles, 22*, GB3009, doi:10.1029/2007GB003050.

Li, K. F., B. Tian, D. E. Waliser, and Y. L. Yung (2010), Tropical mid-tropospheric CO2 variability driven by the Madden-Julian oscillation, *PNAS, 107*(45), 19171- 19175, doi:10.1073/pnas.1008222107.

Le Quéré, C. et al. (2013), The global carbon budget 1959–2011, *Earth Syst. Sci. Data, 5*, 165-185, doi:10.5194/essd-5-165-2013.

Lin, S.-J. (2004), A "vertically Lagrangian" finite-volume dynamical core for global models, *Monthly Weather Review*, *132*(10):2293-2307.

Liu, J., I. Fung, E. Kalnay, J.-S. Kang, E. T. Olsen, and L. Chen (2012), Simultaneous assimilation of AIRS Xco2 and meteorological observations in a carbon climate model with an ensemble Kalman filter, *Journal of Geophysical Research, 117*, D05309, doi:10.1029/2011JD016642.

Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith (2000), A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests, *Monthly Weather Rev*iew, 128, 3187–3199.

Los, S. O, G. J, Collatz, P. J. Sellers, C. M. Malmstrom, N. H. Pllack, R. S. DeFries L. Bounoua, M. T. Parris, C. J. Tucker, and D. A. Dazlich (2000), A global 9-yr biophysical land surface dataset from NOAA AVHRR data, *Journal of Hydrometeorology, 1*, 183-199.

Los, S. O. (2013), Analysis of trends in fused AVHRR and MODIS NDVI data for 1982–2006: Indication for a CO_2 fertilization effect in global vegetation, *Global Biogeochemical Cycles*, *27*, 318–330, doi:10.1002/gbc.20027.

Losch, M., D. Menemenlis, P. Heimbach, J. Campin, and C. Hill (2010), On the formulation of sea-ice models. Part 1: Effects of different solver implementations and parameterizations, *Ocean Modelling, 33*, 129–144.

Maksyutov, S., et al. (2013), Regional CO2 flux estimates for 2009–2010 based on GOSAT and ground-based CO2 observations, *Atmospheric Chemistry and Physics, 13*, 9351-9373, doi:10.5194/acp-13-9351-2013.

Mann, G. W., et al. (2012), Intercomparison of modal and sectional aerosol microphysics representations within the same 3-D global chemical transport model, *Atmospheric Chemistry and Physics, 12*, 4449-4476, doi:10.5194/acp-12-4449-2012.

Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey (1997a), A finite-volume, incompressible Navier-Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res., 102*, 5753–5766.

Marshall, J., C. Hill, L. Perelman, and A. Adcroft (1997b), Hydrostatic, quasi-hydrostatic and non-hydrostatic ocean modeling, *Journal of Geophysical Research, 102*, 5733–5752.

Menemenlis, D., C. Hill, A. Adcroft, J. Campin, B. Cheng, B. Ciotti, I. Fukumori, P. Heimbach, C. Henze, A. Koehl, T. Lee, D. Stammer, J. Taft, and J. Zhang (2005a), NASA supercomputer improves prospects for ocean climate research, *Eos Trans. AGU, 86*, 89, 95–96.

Menemenlis, D., I. Fukumori, and T. Lee (2005b), Using Green's functions to calibrate an ocean general circulation model, *Monthly Weather Review, 133*, 1224–1240.

Menemenlis, D., J. Campin, P. Heimbach, C. Hill, T. Lee, A. Nguyen, M. Schodlock, and H. Zhang (2008), ECCO2: High resolution global ocean and sea ice data synthesis, *Mercator Ocean Quarterly Newsletter, 31*, 13–21.

Messerschmidt, J., et al. (2012), Automated ground-based remote sensing measurements of greenhouse gases at the Białystok site in comparison with collocated in situ measurements and model data, *Atmospheric Chemistry and Physics, 12*(15), 6741–6755, doi:10.5194/acp-12-6741-2012.

Moorthi S., and M. J. Suarez (1992), Relaxed Arakawa–Schubert: A parameterization of moist convection for general circulation models, *Monthly Weather Review*, 120, 978–1002.

Mu, M., et al. (2011), Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, *Journal of Geophysical Research,* 116, D24303, doi:10.1029/2011JD016245.

Najjar, R. G., et al. (2007), Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2), *Global Biogeochemical Cycles, 21*, GB3007, doi:10.1029/2006GB002857.

Novelli, P. C., L. P. Steele, and P. P. Tans (1992), Mixing ratios of carbon monoxide in the troposphere, *Journal of Geophysical Research, 97*, 20,731-20,750.

O'Dell, C. W. et al. (2012), The ACOS CO2 retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech., 5*, 99-121, doi:10.5194/amt-5-99-2012.

Olsen, S. C., and J. T. Randerson (2004), Differences between surface and column atmospheric CO2 and implications for carbon cycle research, *Journal of Geophysical Research, 109*, D02301, doi:10.1029/2003JD003968.

Orr, J. C., et al. (2001), Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models, *Global Biogeochem. Cycles*, *15*(1), 43–60, doi:10.1029/2000GB001273.

Ott, L., B. Duncan, S. Pawson, P. Colarco, M. Chin, C. Randles, T. Diehl, and J. E. Nielsen (2010), Influence of the 2006 Indonesian Biomass Burning Aerosols on Tropical Dynamics Studied with the

GEOS-5 AGCM, *Journal of Geophysical Research, 115*. doi:10.1029/2009JD013181.

Ott, L., S. Pawson, and J. Bacmeister (2011), An analysis of the impact of convective parameter sensitivity on simulated global atmospheric CO distributions, *Journal of Geophysical Research, 116*, D21310. doi: 10.1029/2011JD016077.

Piao, S. et al. (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, *Global Change Biology, 19*, 2117–2132. doi: 10.1111/gcb.12187.

Popova, E. E., A. Yool, A. C. Coward, F. Dupont, C. Deal, S. Elliott, E. Hunke, M. Jin, M. Steele, and J. Zhang (2012), What controls primary production in the Arctic Ocean? Results from an intercomparison of five general circulation models with biogeochemistry, *Journal of Geophysical Research, 117*, C00D12, doi:10.1029/2011JC007112.

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

Potter, C., S. Klooster, A. Huete, and V. Genovese (2007), Terrestrial carbon sinks for the United States predicted from MODIS satellite data and ecosystem modeling, *Earth Interactions, 11*, 1-21.

Potter, C., S. Klooster, V. Genovese, C. Hiatt, S. Boriah, V. Kumar, V. Mithal, and A. Garg (2012), Terrestrial ecosystem carbon fluxes predicted from MODIS satellite data and large-scale disturbance modeling, *International Journal of Geosciences*, doi:10.4236/ijg.2012.

Raczka, B. M. et al. (2013), Evaluation of continental carbon cycle simulations with North American flux tower observations, *Ecological Monographs, 83*:4, 531-556.

Randerson J. T., M. V. Thompson, and C. M. Malmstrom (1996), Substrate limitations for heterotrophs: Implications for models that estimate the seasonal cycle of atmospheric CO2, *Global Biogeochemical Cycles, 10*, 585-602.

Randerson, J. T. G. J. Collatz, J. E. Fessenden, A. D. Munoz, C. J. Still, J. A. Berry, I. Y. Fung, N. Suits, and A. S. Denning (2002), A possible global covariance between terrestrial gross primary production and 13C discrimination: Consequences for the atmospheric 13C budget and its response to ENSO, *Global Biogeochemical Cycles, 16*, doi: 10.1029/2001GB001845.

Randerson, J. T. et al. (2009), Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models, *Global Change Biology, 15*, 2462–2484. doi: 10.1111/j.1365-2486.2009.01912.

Reinecker, M. M., Suarez, M. J., Todling, R., et al. (2008), The GEOS-5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0, Tech. Rep. 104606 V27, NASA.

Rienecker, M. M., et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *Journal of Climate, 24*, 3624–3648.

Ruzmaikin, A., H. H. Aumann, and T. S. Pagano (2012), Patterns of CO2 Variability from Global Satellite Data, *Journal of Climate*, doi: 10.1175/JCLI-D-11-00223.1

Schaefer, K., et al. (2012), A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, *Journal of Geophysical Research*, *117,* G03010, doi: 10.1029/2012JG001960.

Schwalm, C. R. et al. (2010), A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis, *Journal of Geophysical Research, 115*, G00H05, doi:10.1029/2009JG001229.

Sentman, L. T., E. Shevliakova, R. J. Stouffer, and S. Malyshev (2011), Time Scales of Terrestrial Carbon Response Related to Land-Use Application: Implications for Initializing an Earth System Model. *Earth Interact.*, 15, 1–16. doi: http://dx.doi.org/10.1175/2011EI401.1.

Takagi, H., et al. (2011), On the Benefit of GOSAT Observations to the Estimation of Regional CO2 Fluxes, *SOLA*, 2011, Vol. 7, 161-164, doi:10.2151/sola.2011-041.

Takahashi, T., et al. (2002), Global sea-air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects, *Deep-Sea Res. II*, *49,* 1601-1622.

Takahashi, T., et al. (2009). Climatological mean and decadal changes in surface ocean pCO2, and net sea-air CO2 flux over the global oceans, *Deep-Sea Res. II*, *56*, 554-577.

Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric CO2 budget, *Science, 247*, 1431-1438.

Tucker, C. J., J. E. Pinzon, M. E. Brown, D. A. Slayback, E. W. Pak, R. Mahoney, E. F. Vermote, and N. El Saleous (2005), An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *International Journal of Remote Sensing*, 26(20), 4485-4498.

van der Werf, G. R., J. T. Randerson, G. J. Collatz, L. Giglio, P. S. Kasibhatla, A. F. Arellano, S. C. Olsen, and E. S. Kasischke (2004), Continental-scale partitioning of fire emissions during the 1997 to 2001 El Nino/La Nina period, *Science, 303*, 73-76.

van der Werf G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and A. F. Arellano (2006), Interannual variability of global biomass burning emissions from 1997 to 2004, *Atmospheric Chemistry and Physics, 6*, 3423-3441.

van der Werf G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. DeFries, Y. Jin, and T. T. van Leeuwen (2010), Global fire emissions and the contribution of deforestation, agriculture, and peat fires (1997-2009), *Atmospheric Chemistry and Physics, 10*, 11707-11735.

Vanuytrecht, E., D. Raes, and P. Willems (2011), Considering sink strength to model crop production under elevated atmospheric CO2, *Agricultural and Forest Meteorology, 151*, 1753-1762, ISSN 0168-1923, http://dx.doi.org/10.1016/j.agrformet.2011.07.011.

Wang, H., D. J. Jacob, M. Kopacz, D. B. A. Jones, P. Suntharalingam, J. A. Fisher, R. Nassar, and S. Pawson (2009), Error Correlation Between CO2 and CO as a Constraint for CO2 Flux Inversions Using Satellite Data, *Atmos. Chem. Phys., 9*, 7313-7323. doi:10.5194/acp-9-7313-2009.

Wanninkhof, R. (1992), Relationship between wind speed and gas exchange over the ocean, *J. Geophys. Res*., 97(C5), 7373–7382.

Winderlich, J., C. Gerbig, O. Kolle, and M. Heimann (2014), Inferences from CO_2 and CH_4 concentration profiles at the Zotino Tall Tower Observatory (ZOTTO) on regional summertime ecosystem fluxes, Biogeosciences, 11, 2055-2068, doi:10.5194/bg-11-2055-2014.

Wofsy, S. C., et al. (2011), HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, *Proc. Roy. Soc. A, 369*, 2073-2086, doi:10.1098/rsta.2010.0313.

Wunch, D., et al. (2010), Calibration of the Total Carbon Column Observing Network using aircraft profile data, *Atmospheric Measurement Techniques, 3*(5), 1351-1362, doi:10.5194/amt-3-1351-2010.

Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011a), The total carbon column observing network, *Philosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences, 369*(1943), 2087-2112, doi:10.1098/rsta.2010.0240.

Wunch, D., et al. (2011b), A method for evaluating bias in global measurements of CO2 total columns from space, *Atmospheric Chemistry and Physics, 11(23)*, 12317-12337, doi:10.5194/acp-11-12317-2011.

Wunsch, C. and P. Heimbach (2007), Practical global ocean state estimation, *Physica D, 230*, 197–208.

Yokota, T., Y. Yoshida, N. Eguchi, Y. Ota, T. Tanaka, H. Watanabe, and S. Maksyutov (2009), Global Concentrations of CO2 and CH4 Retrieved from GOSAT: First Preliminary Results, *SOLA*, 5, 160-163, 2009.

Zhu, Y. and R. Gelaro (2008), Observation Sensitivity Calculations Using the Adjoint of the Gridpoint Statistical Interpolation (GSI) Analysis System, *Mon. Wea. Rev*., 136, 335-351.

Tables

Table 1. Combinations of surface (land biosphere and oceanic) fluxes used in the model computations in the Flux-Pilot Project. Fossil fuel fluxes were identical in all four combinations.

Flux Combination	**Land Biosphere**	**Ocean**	**Total Land and Ocean Flux (Pg C yr^{-1})**	**Annual Mean Growth Rate at MBL sites (ppmv yr^{-1})**
CG-NO	CASA/GFED-3	NOBM	-3.4 (2009), -2.8 (2010)	2.7
AC-NO	Ames CASA	NOBM	-1.2 (2009), -1.9 (2010)	3.5
CG-ED	CASA/GFED-3	ECCO2-Darwin	-2.1 (2009), -1.4 (2010)	3.3
AC-ED	Ames CASA	ECCO2-Darwin	0.1 (2009), -0.5 (2010)	4.1

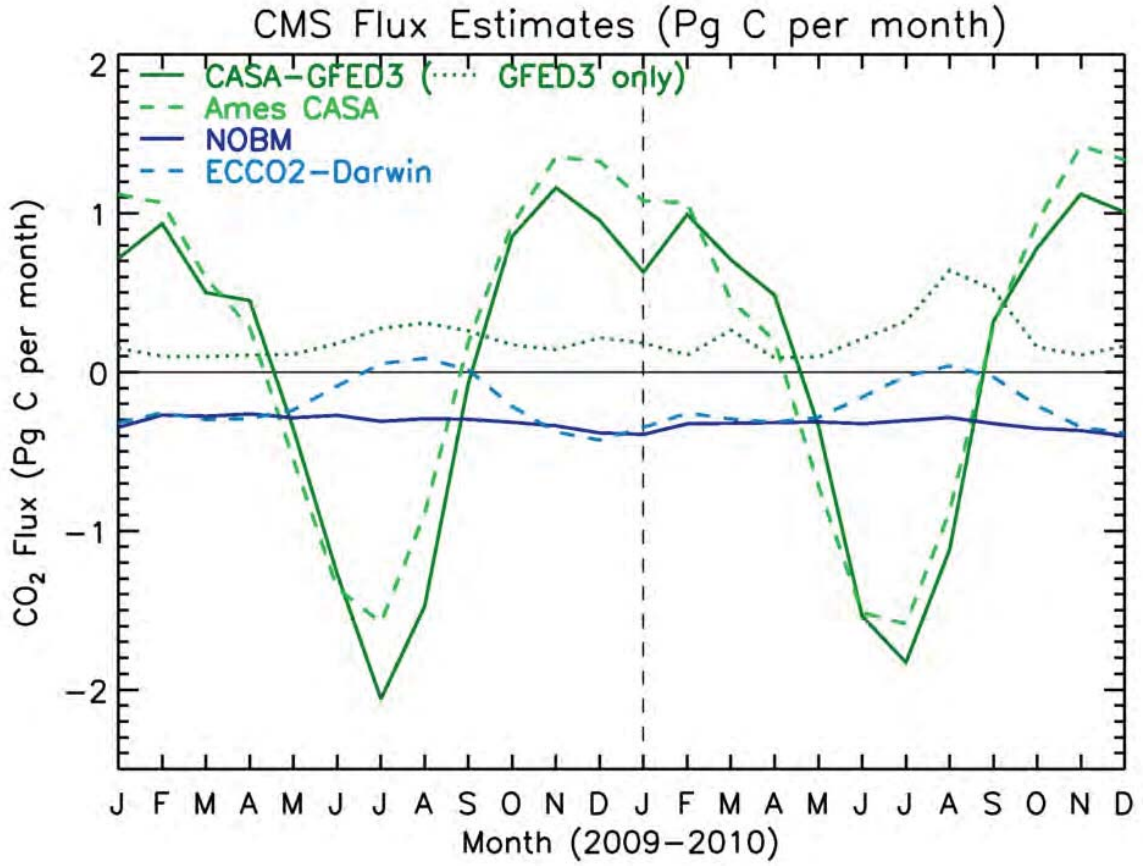


Figure 1. Global land and ocean flux estimates produced as part of NASA's CMS FPP for 2009 and 2010 in Pg carbon per month. Dark green solid line indicates combined NEP and biomass burning fluxes computed from by the CASA-GFED3 model (dotted dark green line shows biomass burning contribution only). Light green dashed line indicates NEP from the Ames CASA model. Blue lines show ocean flux estimates from NOBM (dark blue solid) and ECCO2-Darwin (light blue dashed).

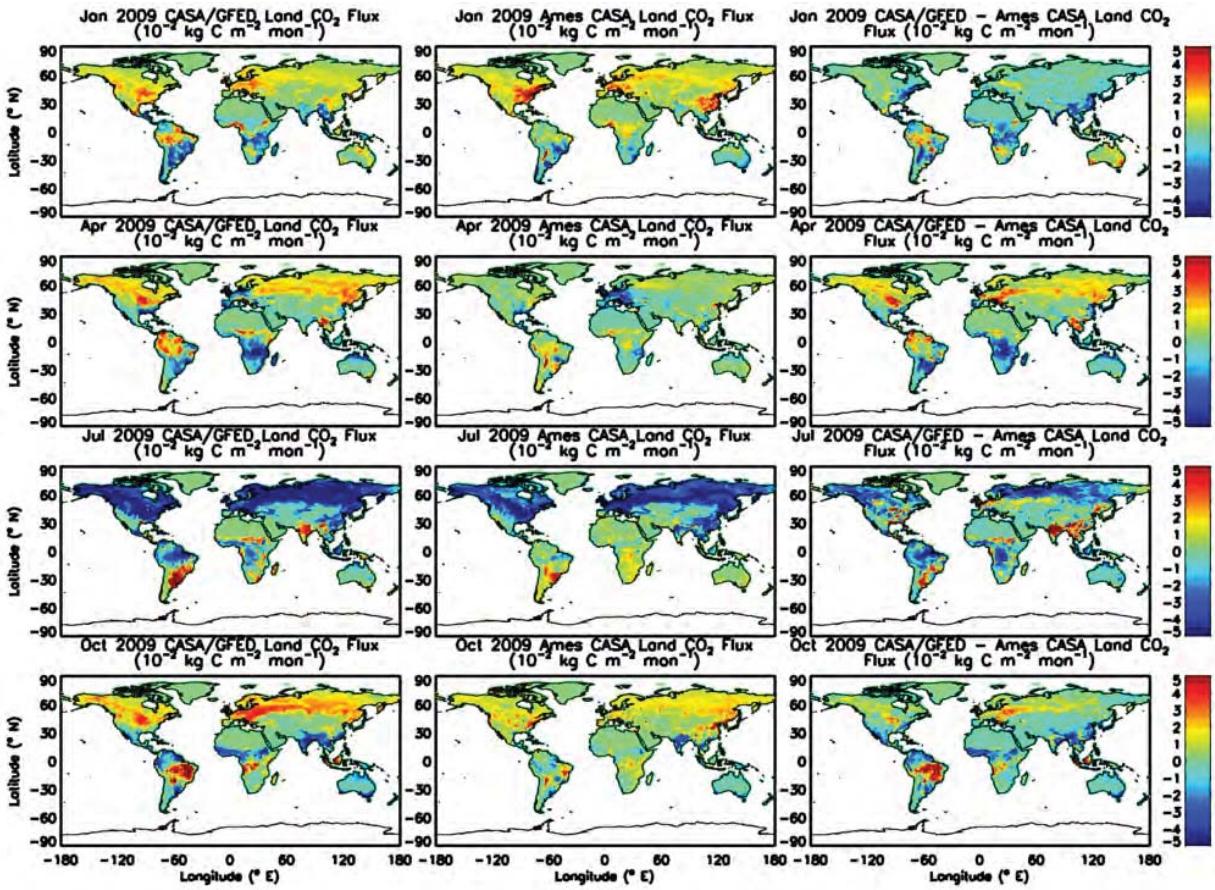


Figure 2. Monthly land flux estimates from the CASA-GFED3 (left) and Ames CASA (middle) models and their difference (CASA-GFED3 minus Ames CASA; right) for January, April, July and October of 2009. Units are 10^{-2} kg Carbon per m^2 per month.

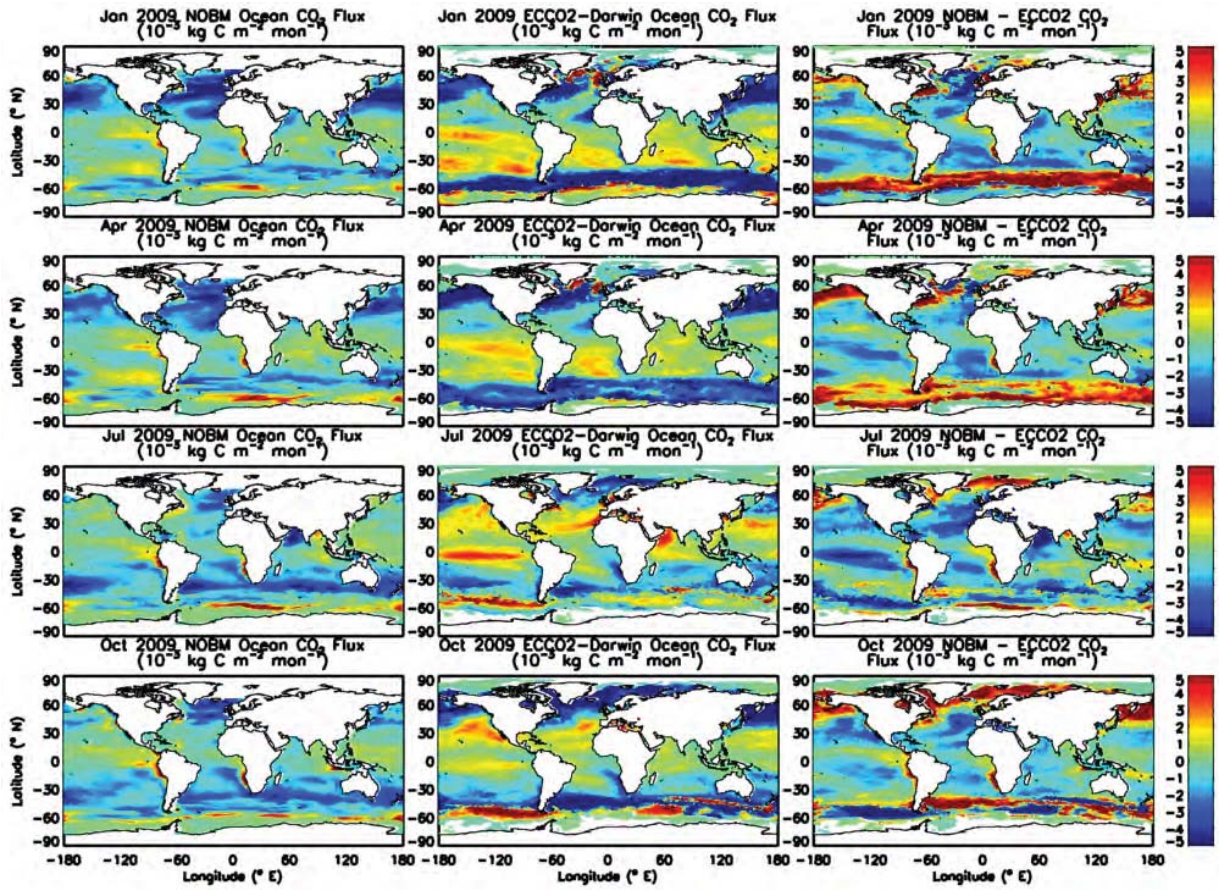


Figure 3. Monthly ocean flux estimates from the NOBM (left) and ECCO2-Darwin (middle) models and their difference (NOBM minus ECCO2-Darwin; right) for January, April, July and October of 2009. Units are 10^{-3} kg Carbon per m^2 per month (note that the ocean fluxes shown here are an order of magnitude smaller than the land fluxes shown in Figure 2).

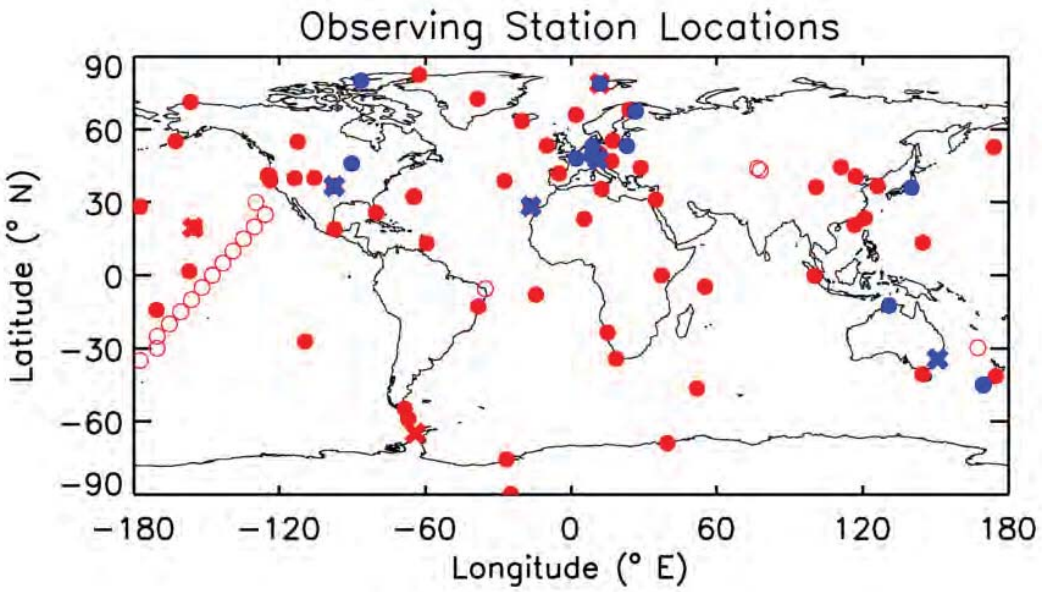


Figure 4. Locations of NOAA (red) and TCCON (blue) observing stations in operation during the study period. Open red circles indicate stations that collected an insufficient amount of data in 2009-2010 and thus were not included in Figures 6-8. Red (blue) 'x' marks indicate NOAA (TCCON) stations shown in Figure 5 (12).

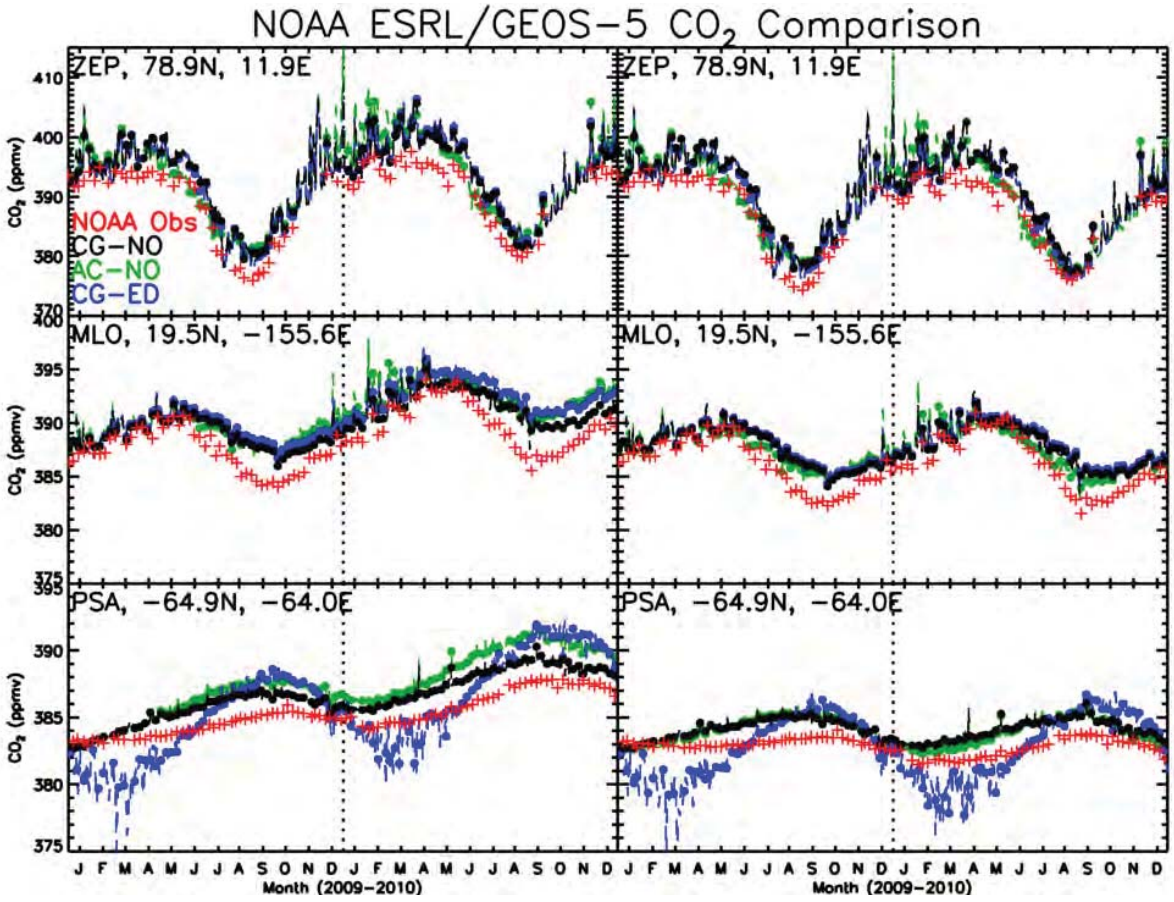
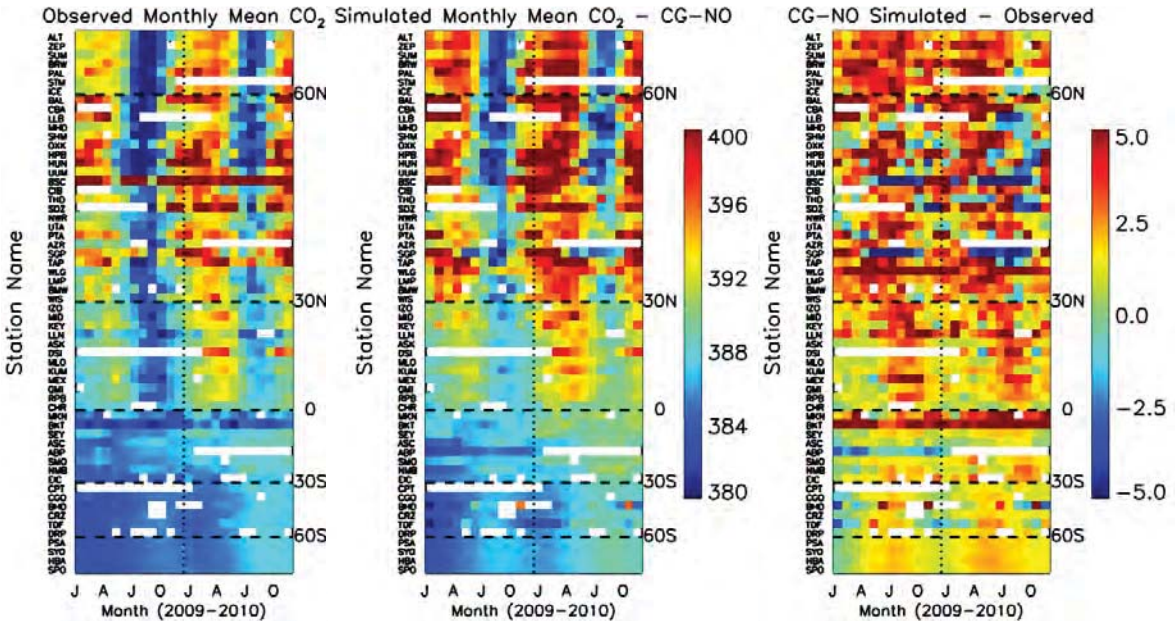


Figure 5. Left panels show simulated and observed CO_2 at the Mauna Loa, Ny-Alesund, and Palmer Station. Right panels show simulated and observed CO_2 at the same locations when differences in atmospheric growth rate are removed. Red indicates NOAA ESRL observations while black, green, and blue indicates GEOS-5 simulated mixing ratios assuming flux scenarios CG-NO, AC-NO, and CG-ED, respectively (flux scenarios are described in detail in Table 1).

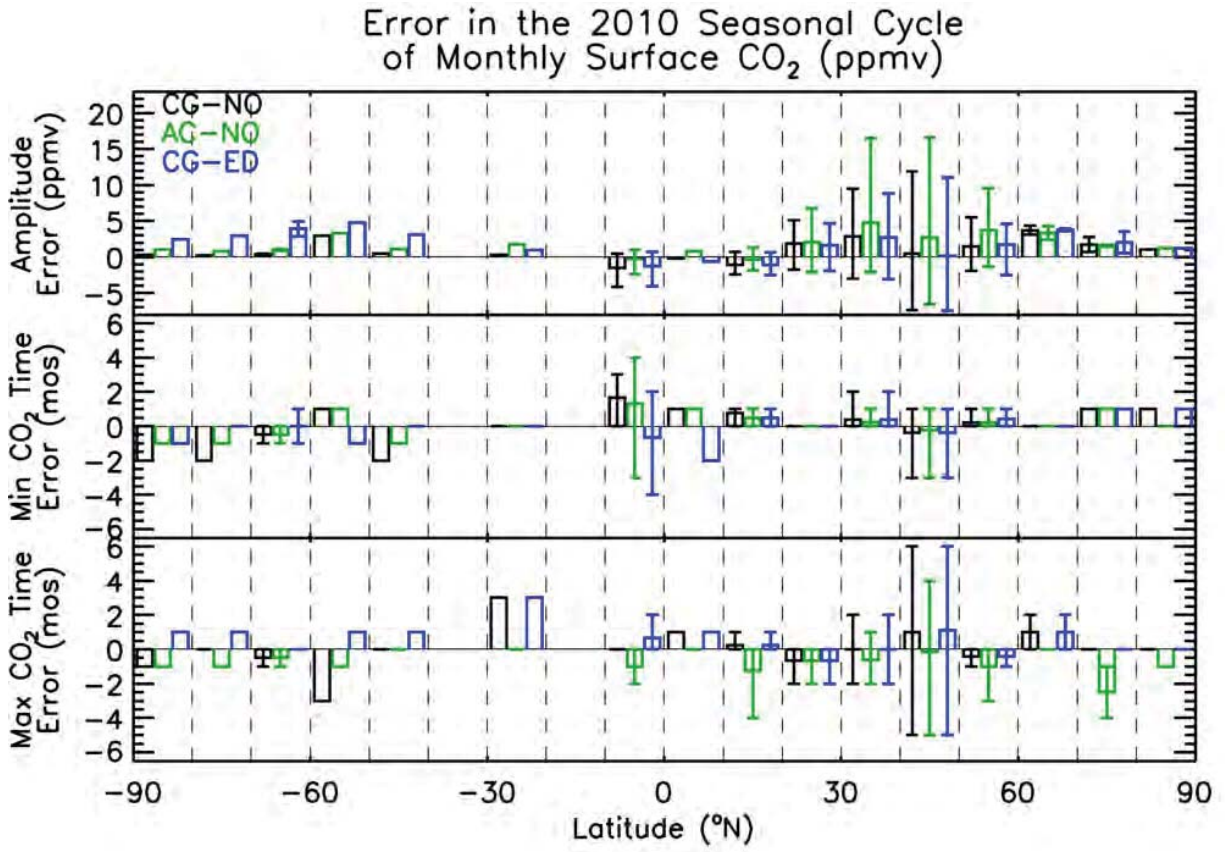


1403

1404 **Figure 6**. Monthly mean CO_2 (ppmv) observed at NOAA ESRL stations during 2009-2010 (left)
1405 compared with simulated CO_2 assuming flux combination CG-NO (middle). Right hand plot shows the
1406 model-observation difference. Stations are oriented from north (top) to south (bottom) on all plots.

1407

1408



1409

1410 **Figure 7.** Evaluation of the simulations' ability to represent the seasonal cycle of CO_2 observed at
1411 NOAA surface stations during 2010. The top plot shows the mean error in the magnitude of the
1412 seasonal cycle calculated over 10 degree latitude bins and for flux combinations CG-NO, AC-NO, and
1413 CG-ED. The middle (bottom) plot shows the mean error in the month during which the minimum
1414 (maximum) in the seasonal cycle occurs. Boxes identify mean values while vertical lines indicate
1415 maximum and minimum values within each 10 degree latitude range.

1416

1417

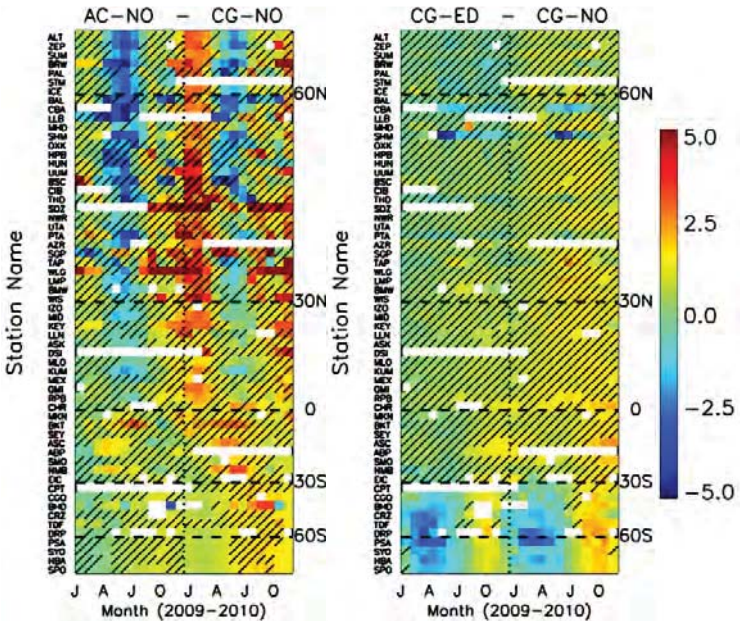
1418

1419

1420

1421

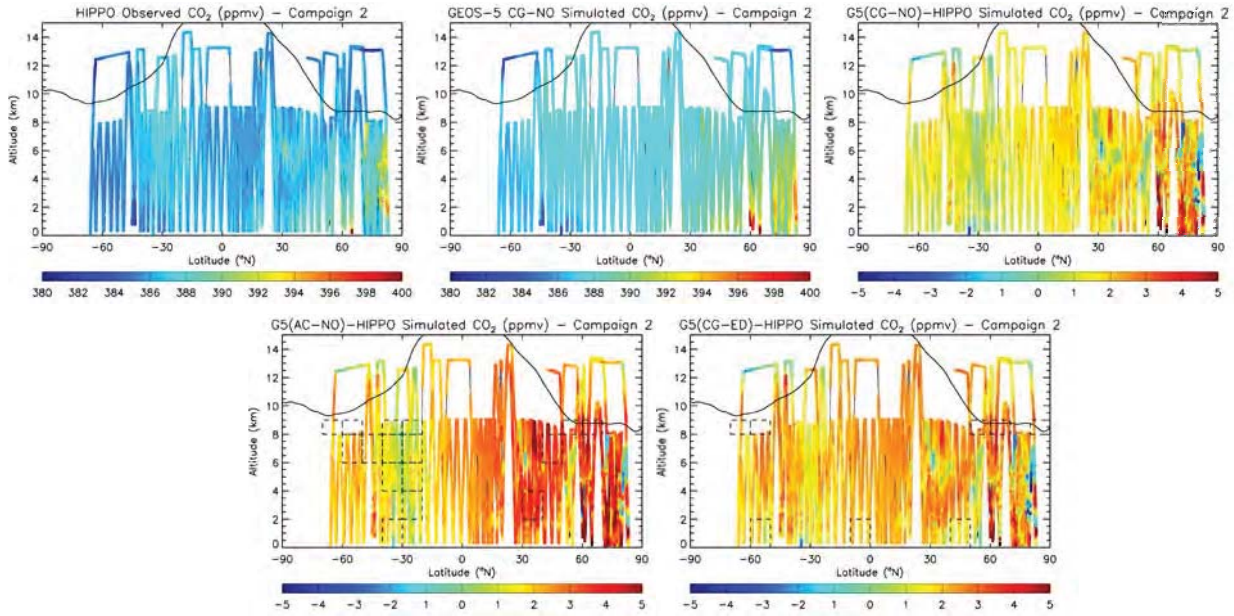
1422



1423

1424 **Figure 8.** Differences in monthly mean CO_2 mixing ratio at NOAA surface stations due to differing
1425 land flux estimates (left) and ocean flux estimates (right). Diagonal bars indicate instances in which
1426 flux differences would not be statistically significant from observation variability.

1427



1432
1433
1434
1435
1436

1437 **Figure 9.** Comparison of observed and simulated CO_2 mixing ratios (top) during HIPPO-2 in October
1438 and November, 2009 show HIPPO observations (left), GEOS-5 using flux combination CG-NO
1439 (middle) and the simulated minus observed difference (right). Bottom plots show the difference
1440 between GEOS-5 and observations using alternate flux scenarios AC-NO (left) and CG-ED (right).
1441 Dashed boxes indicate regions where flux differences are significantly larger than the variance of
1442 observations.

1443

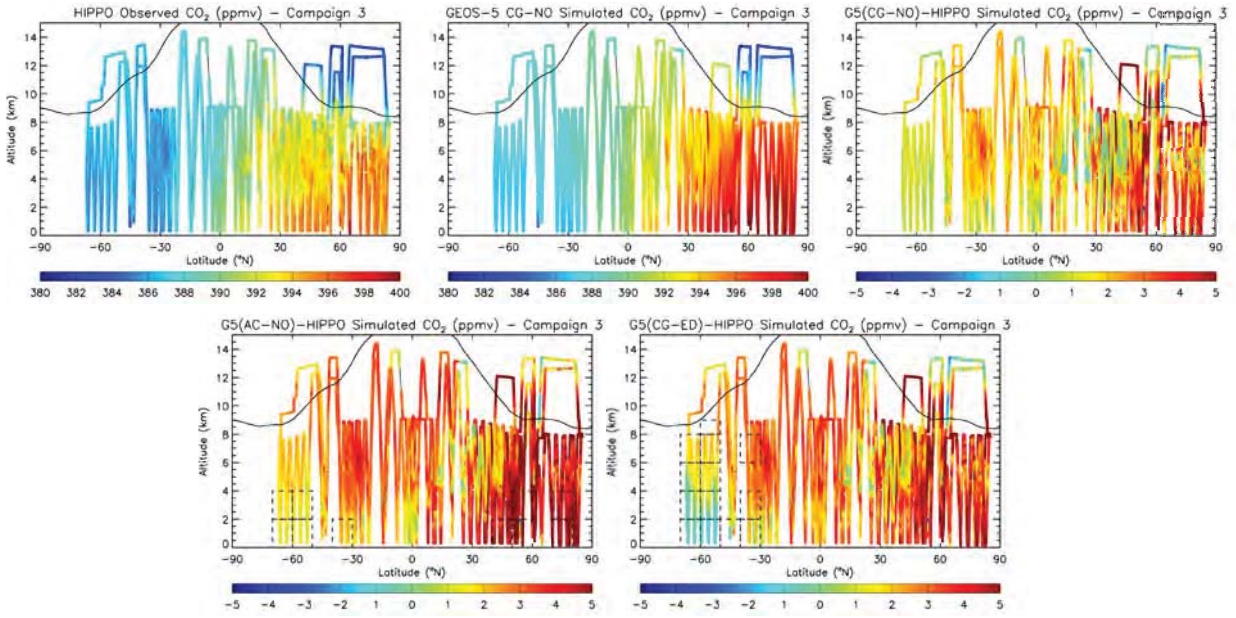
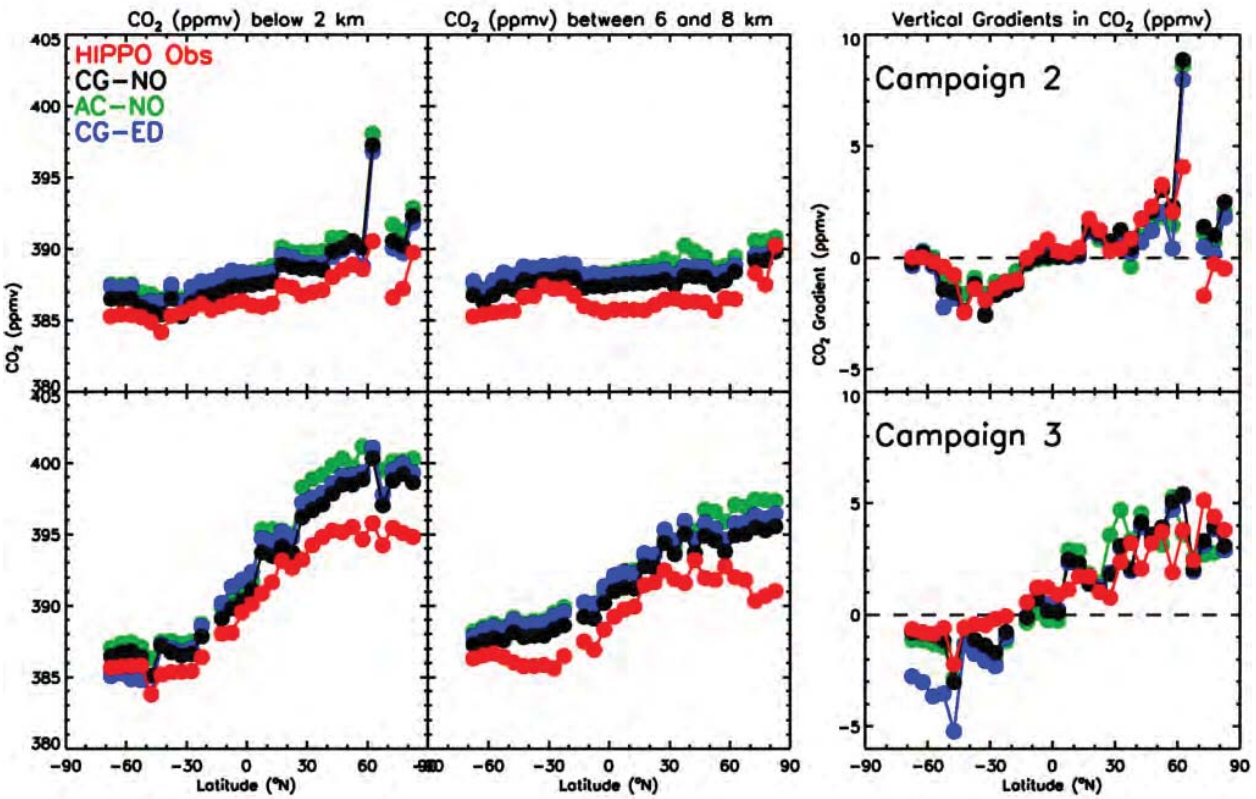


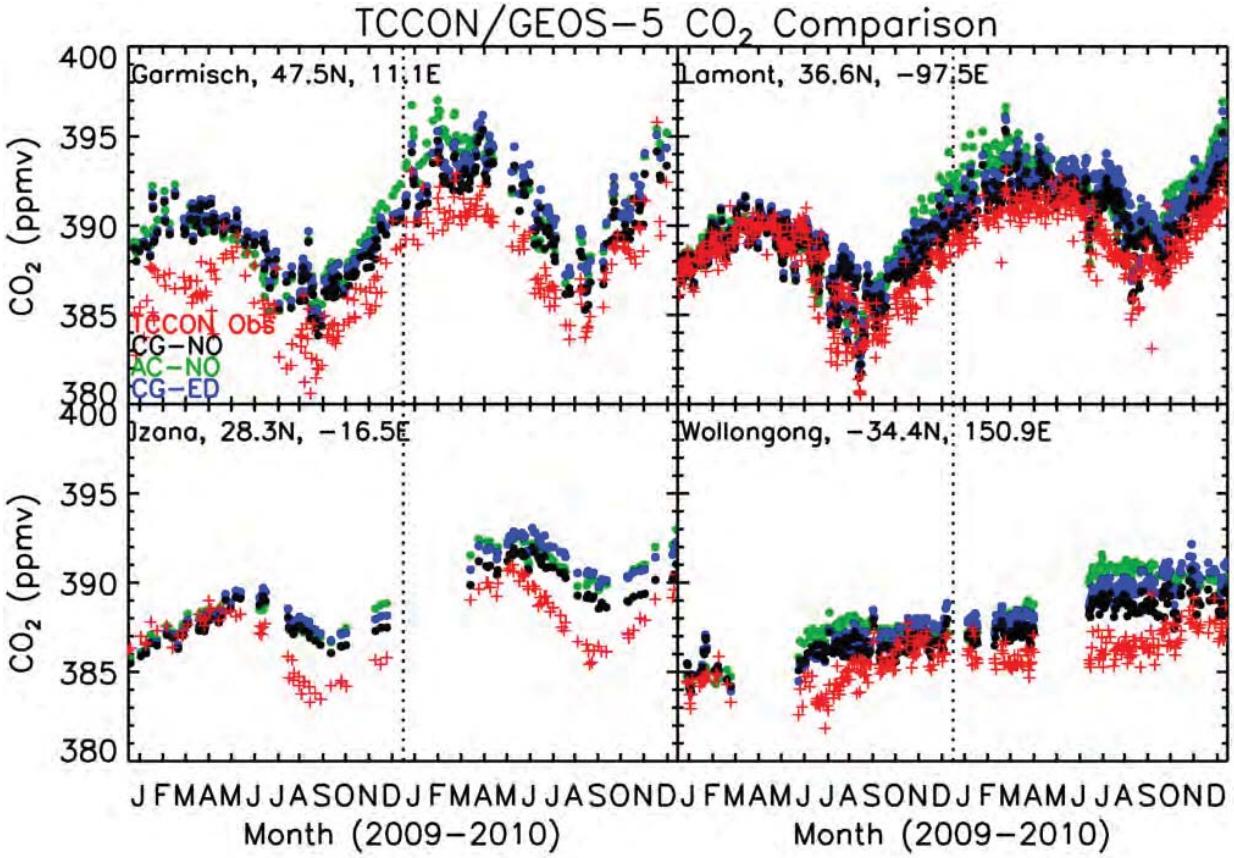
Figure 10. Comparison of observed and simulated CO_2 mixing ratios (top) during HIPPO-3 in March and April, 2010 show HIPPO observations (left), GEOS-5 using flux combination CG-NO (middle) and the simulated minus observed difference (right). Bottom plots show the difference between GEOS-5 and observations using alternate flux scenarios AC-NO (left) and CG-ED (right). Dashed boxes indicate regions where flux differences are significantly larger than the variance of observations.



1459

1460 **Figure 11.** Simulated and observed vertical gradients in in situ CO_2 derived from HIPPO data and
1461 GEOS-5 model simulations (right). Gradients are calculated by binning observations and pseudo-data
1462 into 5 latitude bins and calculating the difference between the mean mixing ratios below 2 km (left) and
1463 between 6 and 8 km (middle).

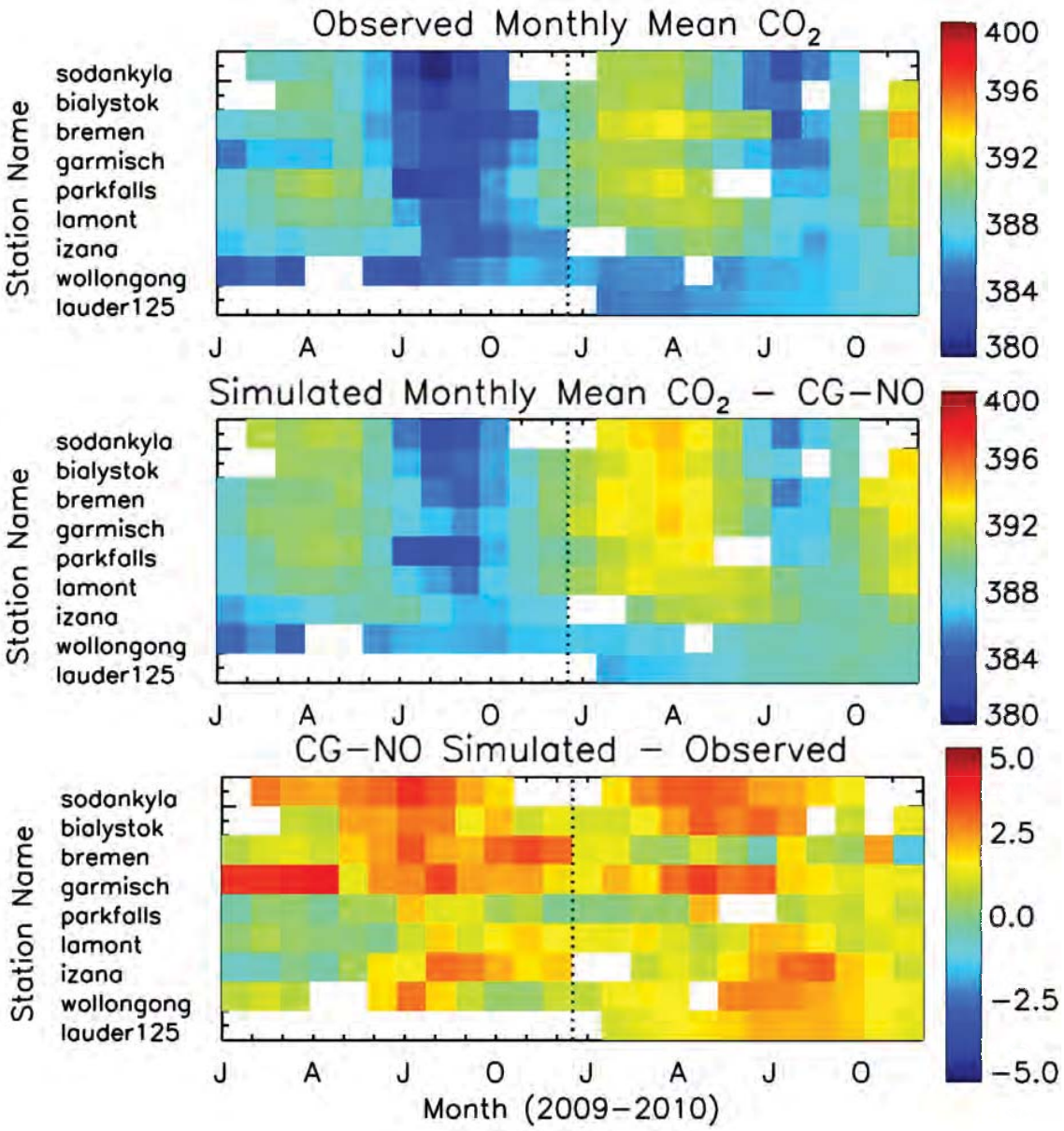
1464



1465

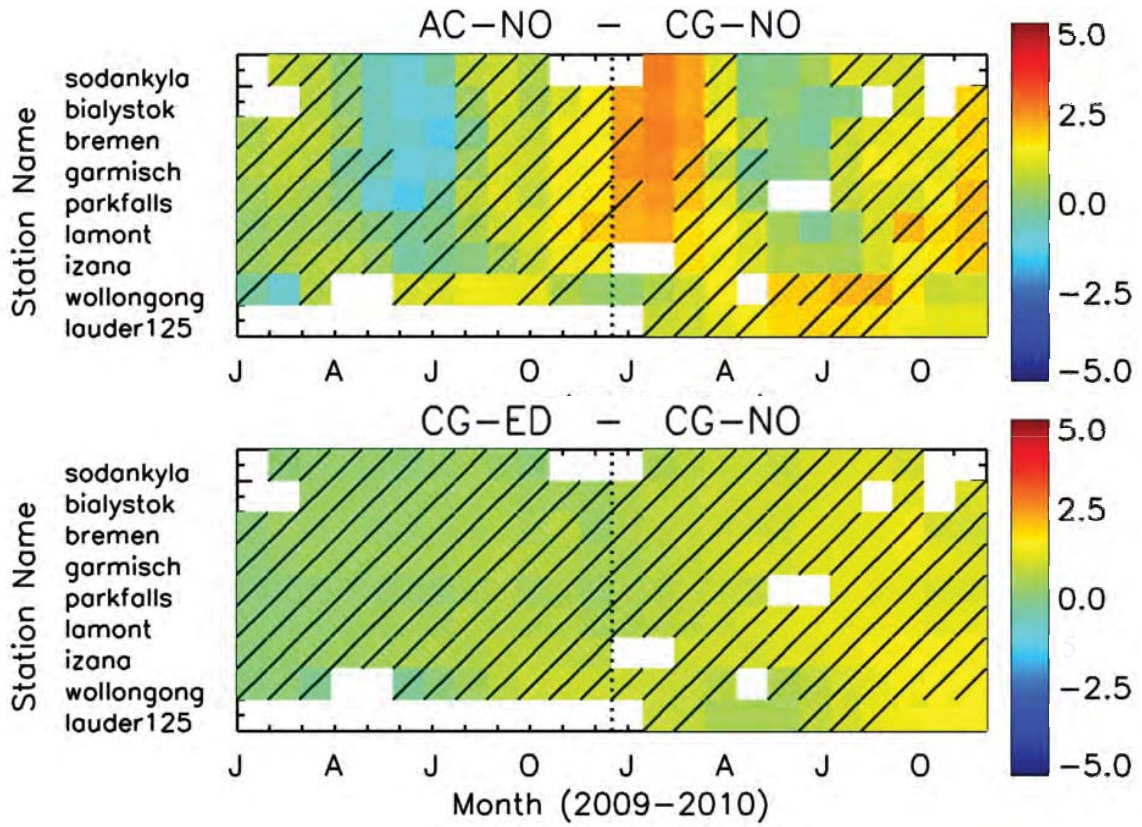
1466 **Figure 12.** Comparison of daily mean simulated and observed column CO_2 at Garmisch, Lamont,
1467 Izana, and Wollongong TCCON stations for 2009-2010.

1468



1487 **Figure 13.** Monthly mean CO_2 (ppmv) observed at TCCON stations during 2009-2010 (top) compared
1488 with simulated CO_2 assuming flux combination CG-NO (middle). Bottom plot shows the model-
1489 observation difference. Stations are oriented from north (top) to south (bottom) on all plots.

1490



1503

1504 **Figure 14.** Differences in monthly mean CO_2 mixing ratio at TCCON stations due to land flux
1505 estimates (top) and ocean flux estimates (bottom). Diagonal bars indicate instances in which flux
1506 differences would not be statistically significant from observation variability.

1507

1508

1509

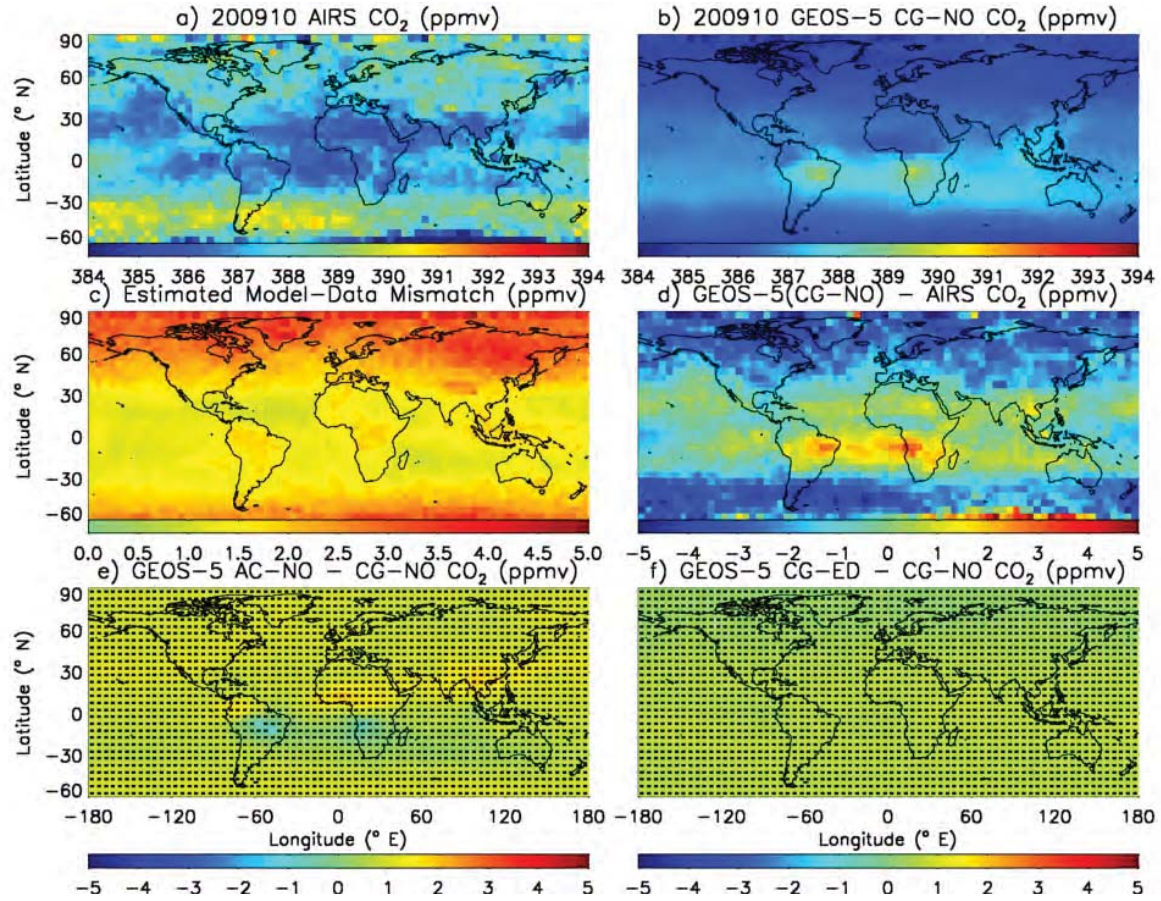


Figure 15. Comparison of AIRS observed mid-troposphere CO_2 mixing ratio with GEOS-5 simulations for October, 2009 for 5 degree latitude by 5 degree longitude grid boxes. The top panels show AIRS monthly mean mixing ratio (left) and monthly mean GEOS-5 mixing ratios assuming flux combination CG-NO sampled using AIRS pressure weighting functions (right). Middle panels show the estimated model data mismatch of AIRS observations for each grid box (left) and GEOS-5 CO_2 minus AIRS (right). Bottom panels show the difference between GEOS-5 simulated CO_2 for flux combinations AC-NO and CG-NO (left) and combinations CG-ED and CG-NO (right); grid cells where these differences are not statistically significant at the 95% confidence level are indicated by black lines.

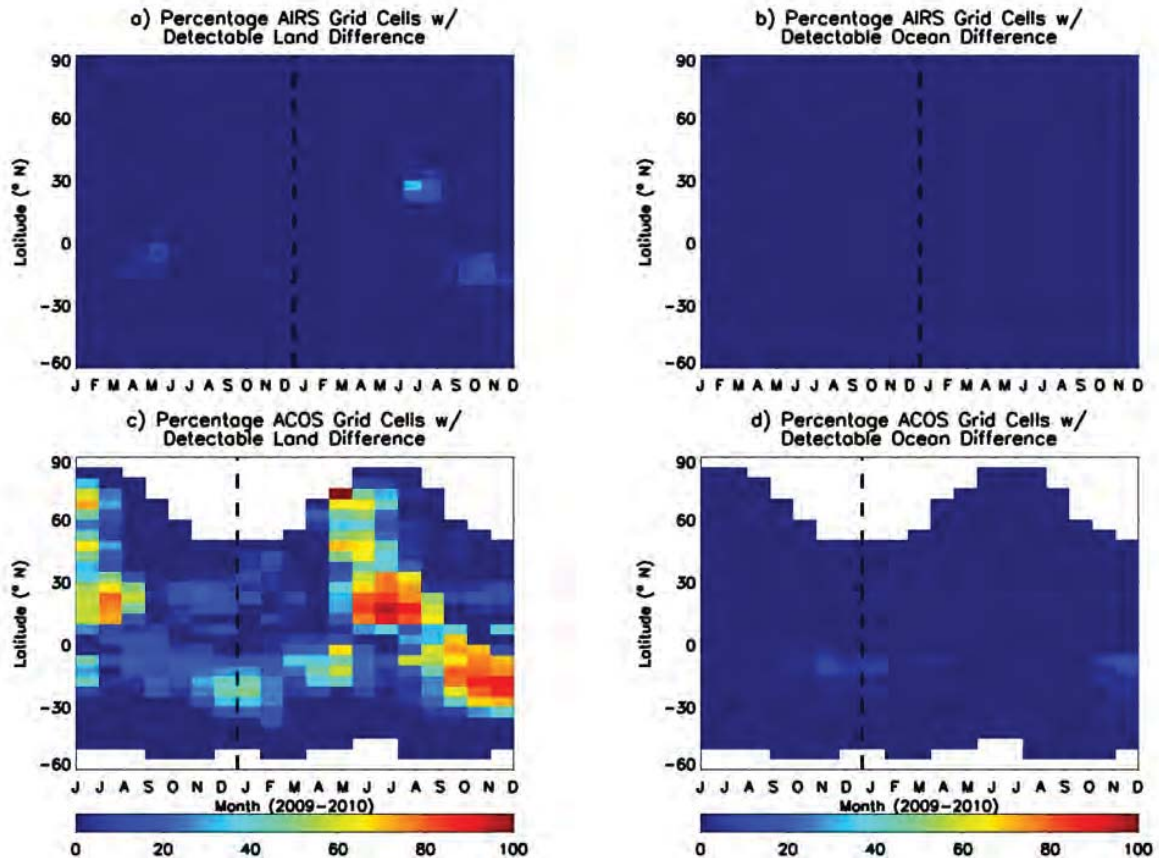
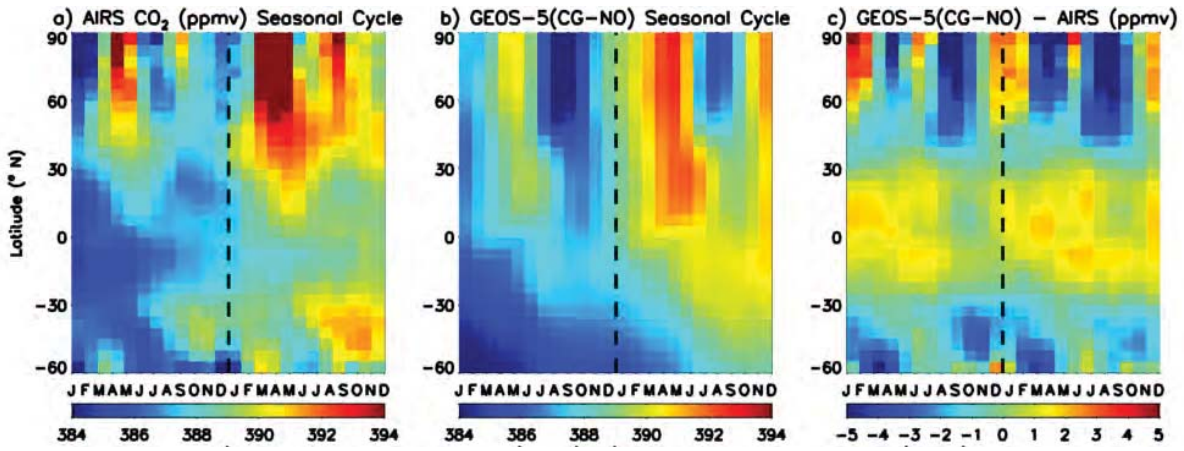


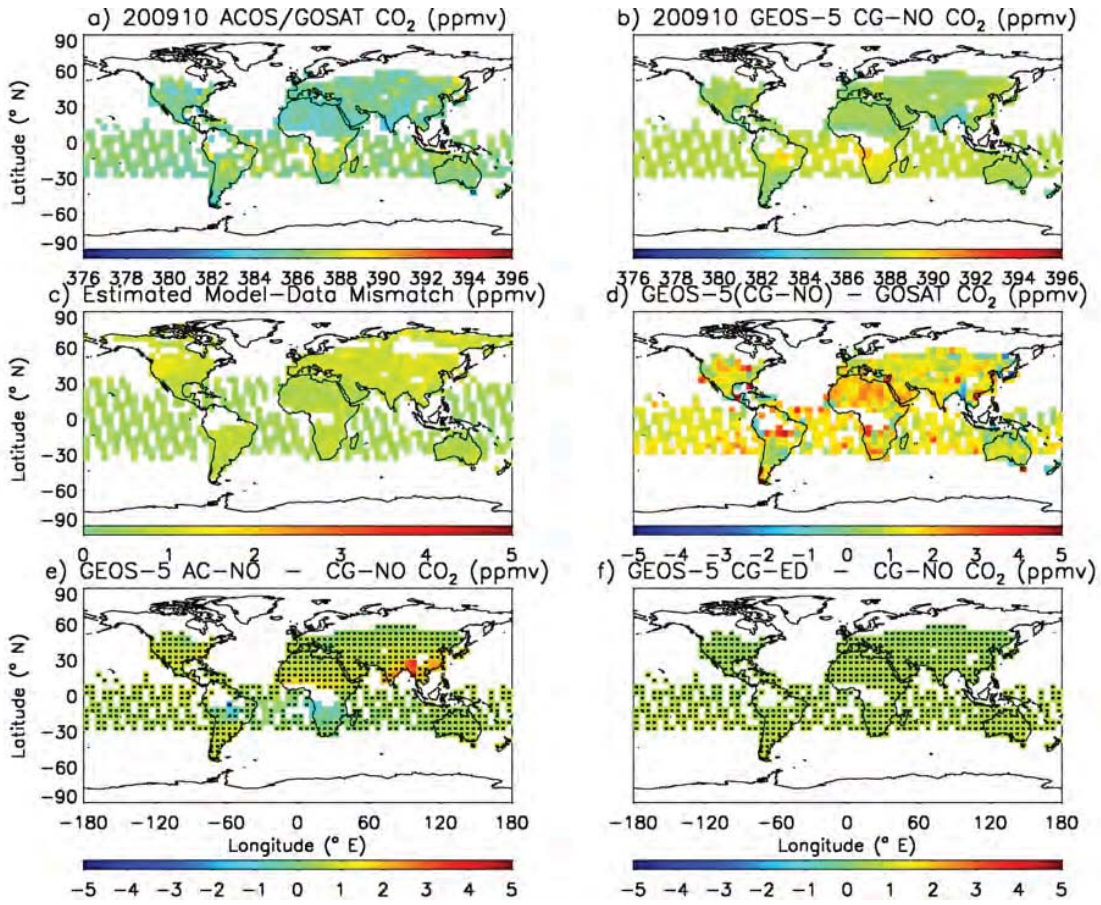
Figure 16. Percentage of 5 degree latitude by 5 degree longitude grid cells per month and per latitude bin in which land (ocean) flux differences manifest as significantly different satellite CO_2 mixing ratios re shown in right (left) panels. Top panels show the detectability of flux differences by AIRS while bottom plots show results for ACOS GOSAT data.



1528

1529 **Figure 17.** Seasonal cycle of (a) zonal mean AIRS observed mixing ratios, (b) zonal mean GEOS-5
1530 (CG-NO) mixing ratios calculated with AIRS weighting functions, and (c) the difference between zonal
1531 mean model results and observations. All units are ppmv.

1532



1533

1534 **Figure 18.** Same as Figure 14, but using ACOS retrievals of GOSAT observations and sampling
1535 instead of AIRS.

1536

1537

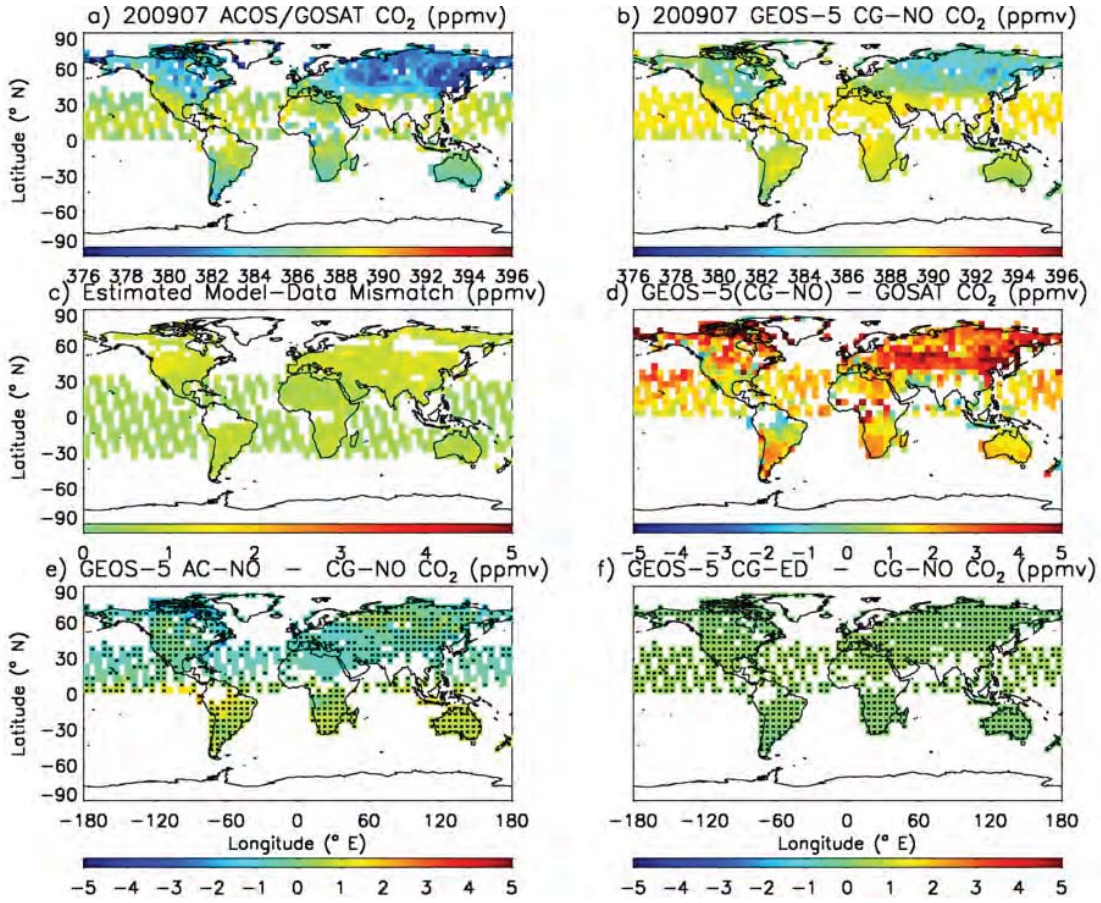
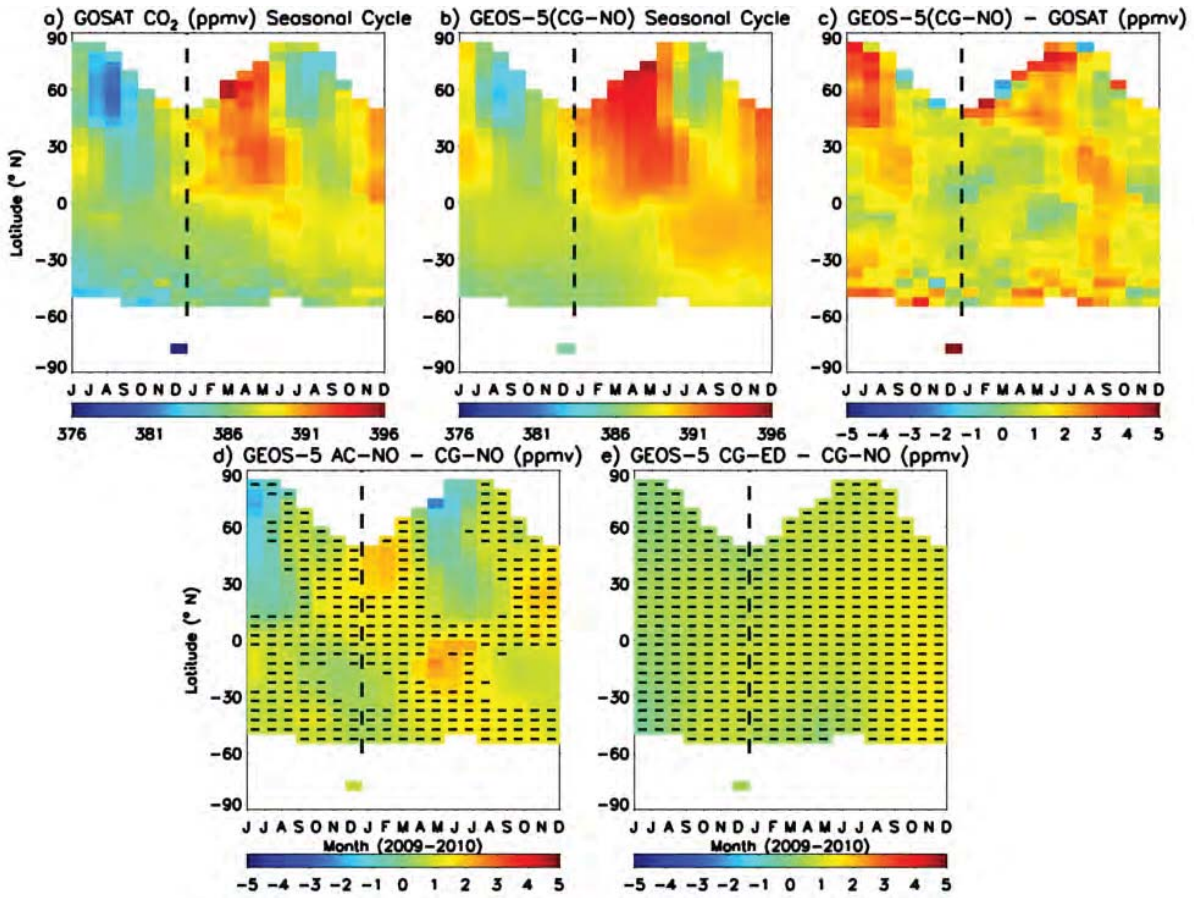


Figure 19. Same as Figure 17, but for July, 2009.



1542

1543 **Figure 20.** Seasonal cycle of (a) zonal mean ACOS observed mixing ratios, (b) zonal mean GEOS-5
1544 (CG-NO) mixing ratios calculated with ACOS weighting functions, and (c) the difference between
1545 zonal mean model results and observations. Bottom plots show the difference in zonal mean caused by
1546 land (d) and ocean flux differences (e). Dashed lines indicate months where flux differences are not
1547 statistically significant from natural variability. All units are ppmv.

1548

1549