Progress Report Year 2: NNX10AT41G
Assimilation, Surface Flux Estimation, and Error Analysis of Atmospheric CO_2 Observations from Space Using a Comprehensive Modeling System

Abstract

Precise estimates of regional sources and sinks of CO2 are necessary for reliable projection of future atmospheric CO2 levels. Top-down estimates of carbon flux from CO2 mixing ratio measurements provide a crucial constraint on the carbon budget partitioned between anthropogenic, oceanic, and biospheric processes. Uncertainties in flux estimates are large due to inadequate observational coverage and the representation of transport relating surface fluxes to observations. While growing observational constraints provided by surface networks and spaced-based platforms have potential to improve the precision of top-down regional flux estimates, uncertainties due to transport are still a major limitation. For example, CO2 variations associated with synoptic weather systems are poorly represented in numerical models and likely to be hidden from satellites by clouds, and may therefore be a source of error for flux inversions. We have performed Observation System Simulation Experiments, or OSSEs, to establish errors in top-down flux estimation due to (1) internal errors in ensemble-based data assimilation and (2) systematic differences in CO2 transport by GEOS4 and GEOS5 analysis products.

We have analyzed GOSAT/ACOS data using a combination of existing models of CO2 exchanges due to hourly photosynthesis and respiration (Baker et al, 2008; Stockli et al, 2008, 2011), crops (Lokupitiya et al, 2009; Corbin et al, 2010), daily air-sea gas exchange (Doney et al, 2009), biomass burning (GFED, Randerson et al, 2007), Fossil Fuel Emissions (Gurney et al, 2009, Corbin et al, 2010), and atmospheric transport (PCTM, Kawa et al, 2004). This comprehensive system allows direct comparison to the observed record of both in-situ and remotely sensed atmospheric CO2 at hourly timescales. We have previously demonstrated that a lower-resolution version of the system has good skill at replicating diurnal, synoptic, and seasonal variations over vegetated land surfaces (Parazoo et al, 2008, 2011). The analysis system will be operated on a 0.5° x 0.67° grid ($\Delta x \sim 50$ km), providing global mesoscale coverage. The system is driven by meteorological output from the NASA Goddard EOS Data Assimilation System, version 5. Surface weather from the system drives calculations of terrestrial ecosystem metabolism (radiation, precipitation, humidity, temperature) and air-sea gas exchange (wind), with other input data coming from satellite data products (e.g., fPAR and LAI from MODIS, and ocean color from SeaWiFS and MODIS).

The result will be estimates of time-varying surface sources and sinks of CO2 that are optimized with respect to in-situ flask and continuous CO2 observations, TCCON data, GOSAT/Tanso retrievals, MODIS data, emissions inventories, and mechanistic models. We will use the modeling and analysis system (1) as a “smart interpolator”

of non-satellite CO2 observations that can be used to estimate systematic errors in GOSAT retrievals; (2) to map and interpret sources and sinks; (3) to quantify the effect of systematic errors in spectroscopic retrievals on source/sink estimates; and (4) to establish detection criteria for fossil fuel emissions.

Project Activities in Year 1

We hired two scientists, purchased a compute cluster for model development and analysis, and conducted end-to-end observing system simulation experiments (OSSEs) in year 1.

Dr. Andrew Schuh is a Research Scientist who has extensive experience in both atmospheric tracer transport and inverse modeling of CO2. He has previously published regional analyses using Lagrangian transport to map sources and sinks of CO2 over North America using in-situ measurements, and compared these results in detail to agricultural and forest inventory data (Schuh et al, 2009). Becky McKeown is a Research Associate who spent a decade in private industry as a software engineer before joining CSU as an ecosystem modeler. She is now finishing a PhD in mathematics working on optimization and data assimilation. Her role in this project is to improve optimization and error propagation methods for satellite data assimilation.

With project support we have procured a mid-sized compute cluster and disk array to perform research tasks, store output, and analyze results. The new compute cluster consist of one head node and three compute nodes. The compute nodes consist of 48 cores and 128 GB of ram each. This give the current cluster configuration a total of 144 cores and 384 GB of ram. The new cluster is used to run large model runs in parallel. To store all the output of the model runs a large disk array is needed. The disk array is configured to run the Solaris operating system in order to take advantage of the ZFS file system. This gives us a very stable filesystem configured with RAID6 to store all the data and the long-term output. This large disk array holds 36 disks given us a usable 60 TB of storage. The array uses RAID6 for data security, which allows the loss of 2 hard disks. The disk storage array is divided up into three arrays of 12 physical disks each. With this large number of disks, RAID6 allows for sufficient time to replace a failed disk and have the raid rebuild without the chance of losing the entire raid while it is rebuilding.

Project Activities in Year 2

In year 2 we have evaluated forward models of fossil fuel emissions, air-sea gas exchange, wildfires, and residual sources and sinks. We have harmonized the various models so that they can all be run on the same global grid using the same weather data (GEOS5), and have used the models as lower boundary forcing for atmospheric transport due to advection by resolved winds, PBL turbulence, and cloud mass flux. Atmospheric transport has been shifted from using PCTM to the closely related (and much better supported) GEOS-Chem, and we have built a system for collecting all input data into the models in near-real time. Finally, we have used the multicomponent modeling system to predict the atmospheric column mean dry-air mole fraction of CO2 and sampled the model at the locations of the actual ACOS/GOSAT retrievals.

Improvements and expansions in software and data required for inversions

The 2010 initial inversions using PCTM and the data sources identified in our 2010 annual report were performed quickly without emphasis on repeatability and validation. There were no established or documented procedures for performing these inversions, and when we experienced some natural turnover in the group we lost a significant amount of system knowledge due to limited overlap in personnel.

In order to prevent a future setback we are implementing and documenting new procedures for data and software management. Our work in 2011 includes software verification, automation of data collection and preprocessing, quality assurance, adherence to portable standard data formats (NetCDF CF/COARDS), and documentation of procedures in order to improve our confidence in results and enhance the accessibility of the data and the inverse framework.

In order to improve model verifiability, we began to consolidate improvements implemented in older versions of PCTM, and Randy Kawa and David Baker's current versions. In previous work multiple versions of PCTM were used: more recent versions for stand-alone forward simulations and an older version that was embedded in the MLEF framework. Comparisons of different versions of PCTM have shown significant, and inexplicable differences in results. Since our goal is to use a transport model, not develop one or maintain it, we began migrating to GeosChem. GeosChem is relatively well documented and maintained, has a broader user base, and a supported database of independently generated fluxes. Every update is tested and benchmarked.

We are in the process of updating the MLEF framework to enable assimilation of GeosChem as well as newer versions of PCTM.

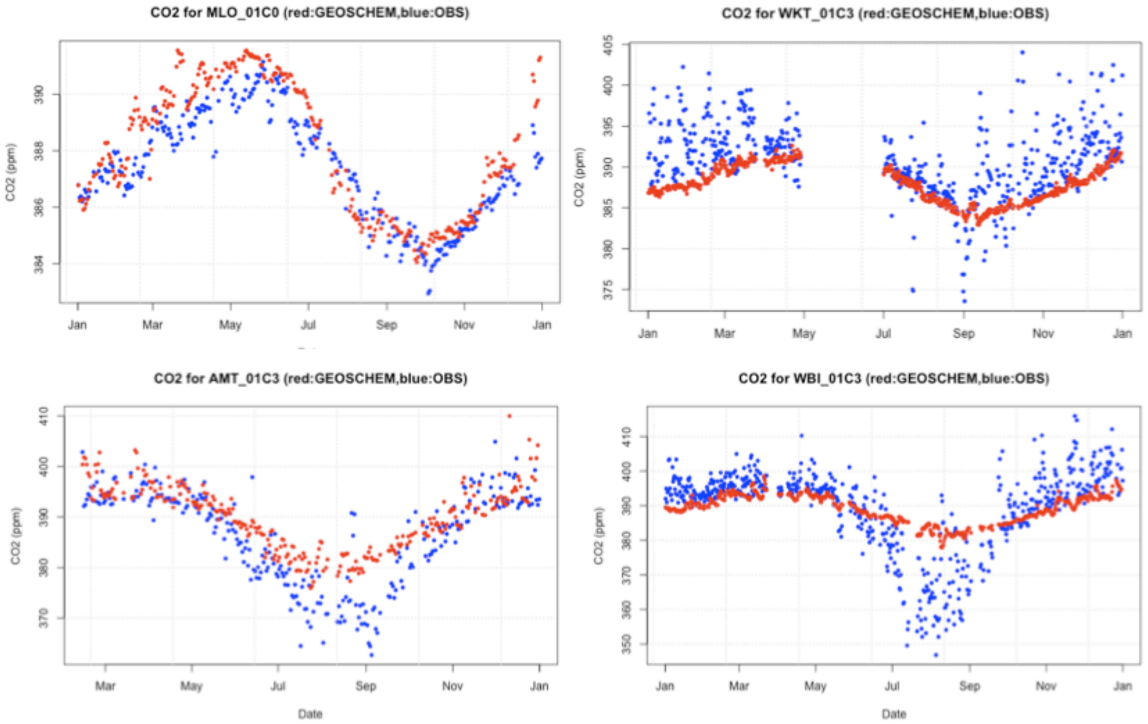
We continued to update flux data for input to the transport model. From Scott Doney, we received daily ocean flux data for 2009 and 2010, as well as consistent monthly ocean flux data for 2000 through 2010 (Doney et al, 2009). We expect to receive consistent daily ocean flux for 2000 through 2010 shortly. In addition, we have acquired the ODIAC high-resolution fossil fuel CO2 emissions (Oda and Maksyutov, 2011). SiB fluxes have been generated using the GEOS5 0.67x0.5 degree dataset using the new prognostic phenology algorithm, but these fluxes have unrealistic, unbalanced respiration over equatorial Africa. We are currently forcing them into balance with the GPP while we attempt to fix the algorithm.

Tools and procedures for pre-processing the flux data for input to PCTM and GeosChem have been developed, archived and documented on our in-house wiki.

We have developed automated download and pre-processing systems for Geos5 and MERRA products to support transport simulations as well as the generation of SiB fluxes.

Comparisons of GEOS-CHEM forward simulations to NOAA surface data and GOSAT

These initial results were obtained using SiB3 fluxes with the phenology defined by MODIS, and ocean and fossil fuel fluxes that are available on the GEOS-CHEM website. GEOS-CHEM (version 9.1.3) was modified in order for 1 hour respiration and gross primary production (GPP) fluxes to be input from the SiB model. These fluxes were created by running the offline SiB model with the same MERRA (GEOS5) driving data as was used in the forward transport simulation. Fossil fuel fluxes from Andres et al 2010 were used as well as ocean fluxes from Takahashi et al. 2009. The code was also modified to input/output netcdf4 files as well as produce customized observation output for both satellite and in-situ observations. Initial forward simulations provide a very good fit to the NOAA surface data on the seasonal cycle. Examples for four arbitrarily chosen sites are shown in Figure *. Note that this version of SiB did not employ the Lokupitiya et al



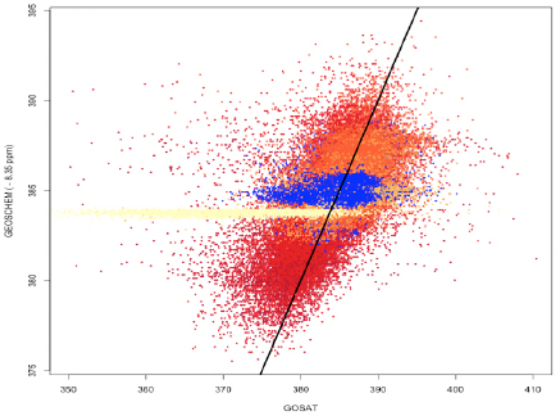


Figure *: GOSAT data plotted against CO2 modeled from GEOSCHEM. Blue dots indicate retrievals over the ocean and "light tan" dots indicate highly uncertain retrievals in

2009 model for crop fluxes and thus the large deviations in CO2 between the model and observations seen at the WBI tower in Iowa.

Comparisons were then made between the simulated CO2 and the latest ACOS GOSAT product. Results showed a strong correlation between the modeled CO2 and ACOS GOSAT product. These results were compared and similar to those from independent runs by David Baker at CIRA. This essentially is a sanity check and provides a degree of confidence in moving forward to both inversion results as well as various comparisons with covariates in the GOSAT retrievals, such as blended albedo and airmass estimates.

We analyzed residuals between the modeled CO2 observations and the GOSAT data for correlations with potential covariates in the GOSAT retrieval algorithm. Weak correlations were found between the “airmass” and “blended albedo” variables coming from the retrievals. Both were on the order of 0.4. Surface pressure did not seem to be correlated with the residuals. The correlation on seasonal time scales (obtained by a 2-3 week moving average) can be seen in Figure *. We will perform additional comparisons such as this one, once inversion results can be obtained for the in-situ data with the MLEF framework.

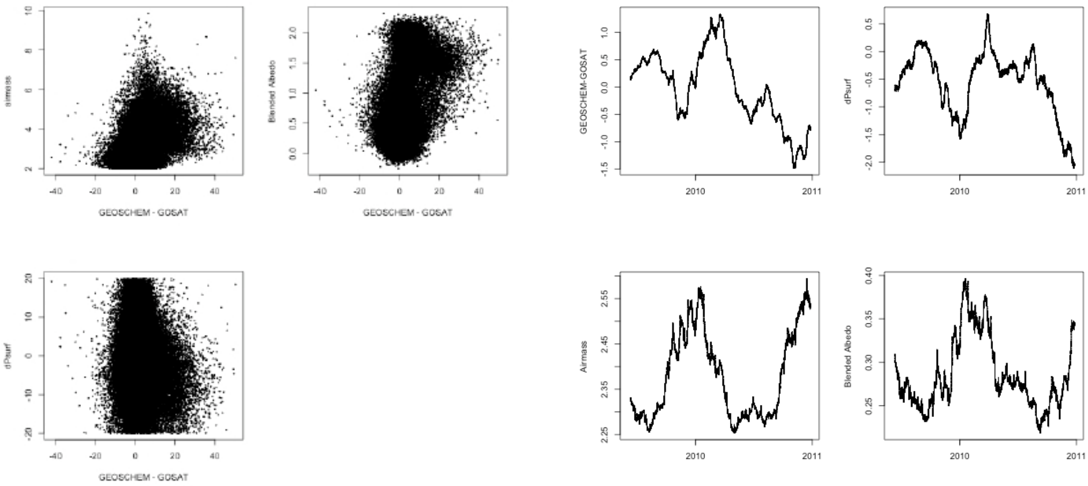
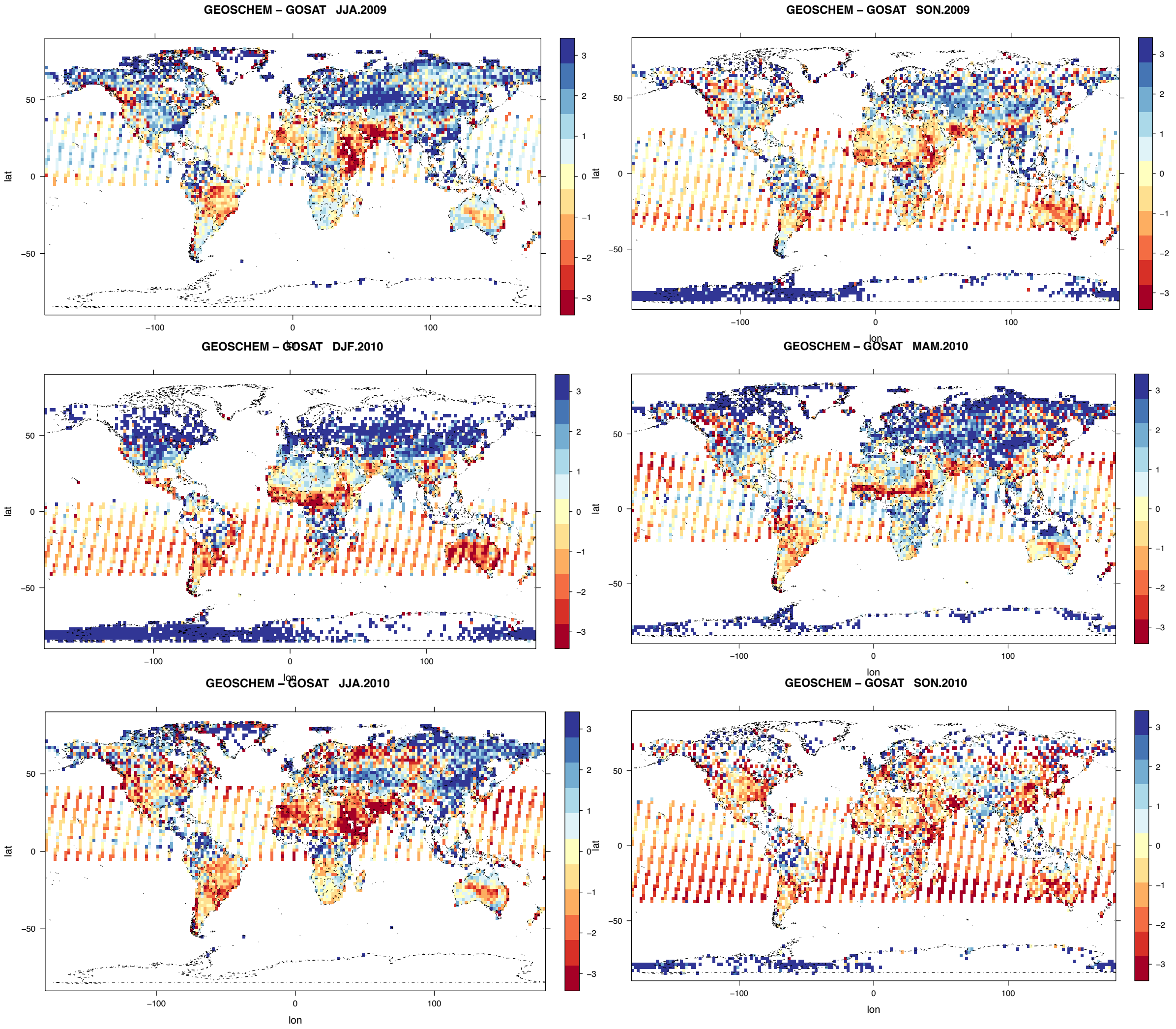


Figure 1: This plot shows the long time scale behavior of the residuals as well as potential explanatory variables. These seem to be the main ones we’re interested in (outside of categorical gain). I used a 10,000 member moving average. Remember this is just “good” data, so you can’t expect the time to be equal between consecutive data records. You might be able to clean this up a bit, but quick and dirty for now. 10,000 represents “roughly” 2-3 week window of data. Reasonable relationship between residuals, blended albedo and airmass. dPsurf not quite as much. Correlations are about 0.4 for the point by point (blend alb and airmass), which seemed high, and higher for this smoothed data you see here. There is a diurnal/latitude effect which is weak but could cause some issues and is currently being smoothed out here.

We further investigated the residuals on a spatial basis, which is not directly possible via simply the residuals. By averaging up the residuals over several 3-month seasons, we were able to identify locations of mean bias (Figure *).



As part of a separate validation exercise for the future ASCENDS lidar satellite mission, we provided forward modeled CO2 differences imposed by realistic, but hypothetically unknown, source-sink structures in the oceans and on the land. In particular, we took the fluxes aforementioned, e.g. annually balanced SiB land fluxes, Takahashi 2009 ocean fluxes, and Andres 2010 fossil fuel fluxes and ran them forward in GEOSCHEM for 2009 and 2010 (control run). We then performed an identical forward simulation with the imposed "flux differences" shown in Figure * (reality run). We allowed 2009 to be a "spin up" year and then looked at the differences between the "reality" and "control" runs as seen through the ASCENDS averaging kernel (2 um). As can be seen in Figure *, certain features such as the strong hypothetical Amazonian sink due to CO2 fertilization can be seen relatively easily throughout the year. Other features such as the reduced respiration effect due to regrowing forests in the Northern midlatitudes can be seen more strongly in the Northern Hemisphere spring and summer.

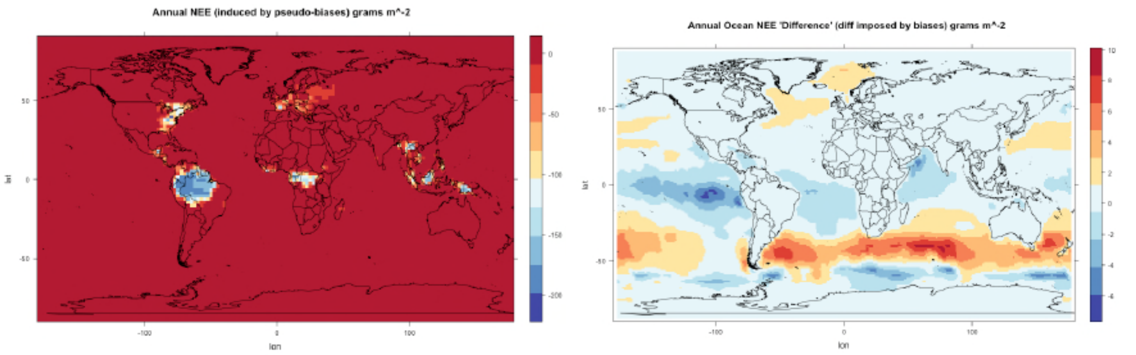


Figure * : Test adjustments to “control” case, sink scenarios over land (left panel) and slower ocean mixing and fluxes (right).

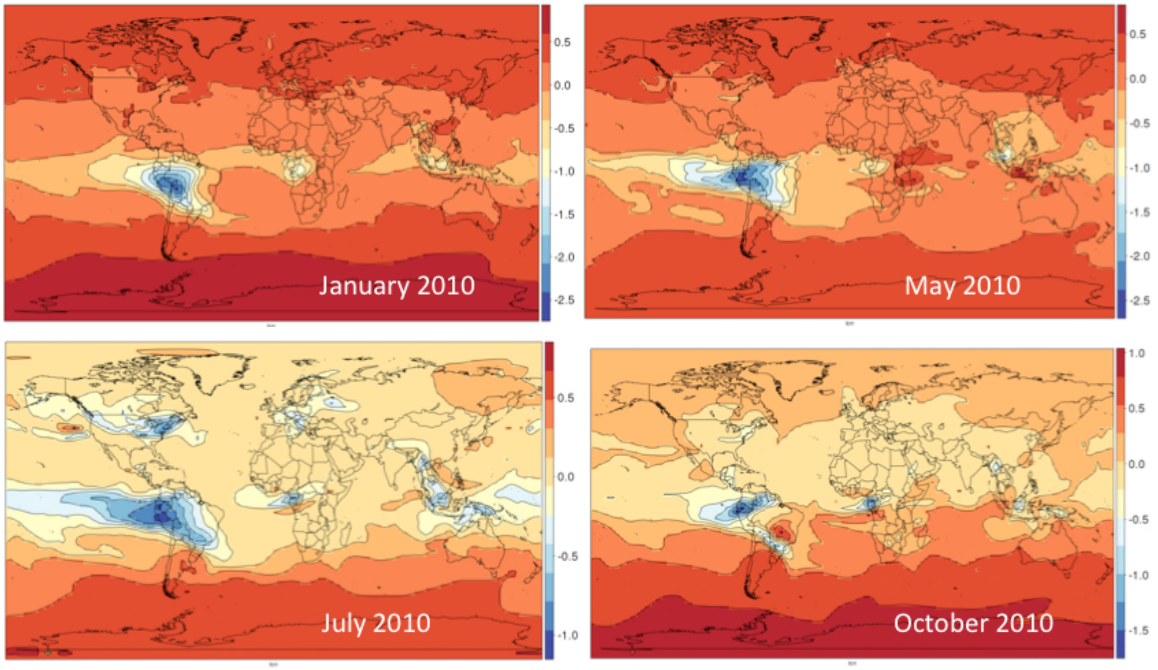


Figure * : Monthly average column differences between “sink” case (figure *) and “control” cases via the 2 um ASCENDS

Schedule of Future Work

- **Year 3:** Build upon the multiyear comparison of the GEOS-CHEM model results to the in-situ observations and GOSAT- TANSO data. Continue with development of the MLEF inversion framework. Analyze in-situ driven inversion results and compare to potential biases in satellite data with respect to surface albedo, aerosol optical depth, and viewing angles. Evaluate results with respect to independent observations not used in the optimization.

References Cited

Baker, I. T., L. Prihodko, A.S. Denning, M. Goulden, S. Miller, and H.R. da Rocha, 2008. Seasonal Drought Stress in the Amazon: Reconciling Models and Observations. *Jour. Geophys. Res.* **113**, G00B01, doi:10.1029/2007JG000644.

Baker, I.T., A.S. Denning, L. Prihodko, K. Schaefer, J.A. Berry, G.J. Collatz, N.S. Suits, R. Stockli, A. Philpott, O. Leonard, 2008: Global Net Ecosystem Exchange (NEE) of CO2, Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Baker, I.T., A. S. Denning, R. Stockli, 2010. North American gross primary productivity: Regional characterization and interannual variability. *Tellus*. **62B**, 533–549. DOI: 10.1111/j.1600-0889.2010.00492.x

Corbin, K.D., A.S. Denning, E.Y. Lokupitya, A.E. Schuh, N.L. Miles, K.J. Davis, S. Richardson, I.T. Baker, 2010: Assessing the impact of crops on regional CO_2 fluxes and atmospheric concentrations. *Tellus*. **62B**, 521–532. DOI: 10.1111/j.1600-0889.2010.00485.x.

Corbin, K. T., A. S Denning, and K. R. Gurney, 2010. The space and time impacts on U.S. regional atmospheric CO2 concentration from a high resolution fossil fuel CO2 emissions inventory. *Tellus*. **62B**, 506-511. DOI: 10.1111/j.1600-0889.2010.00480.x

Doney, S.C., I. Lima, J.K. Moore, K. Lindsay, M.J. Behrenfeld, T.K. Westberry, N. Mahowald, D.M. Glover, and T. Takahashi, 2009b: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, J. Mar. Systems, 76, 95--112.

Gurney, K. R., D. L. Mendoza, Y. Zhou, M. L. Fischer, C. C. Miller, S. Geethakumar and S. de la Rue du Can, 2009. High Resolution Fossil Fuel Combustion CO2 Emission Fluxes for the United States. *Environ. Sci. Technol*., DOI: 10.1021/es900806c

Harper, A. B., A. S. Denning , I. Baker , M. Branson , L. Prihodko , D. Randall, 2010. The role of deep soil moisture in modulating climate in the Amazon rainforest. *Geophys. Res. Lett.*, 2009GL042302.

Kawa, S. R., D. J. Erickson III, S. Pawson, and Z. Zhu (2004), Global CO2 transport simulations using meteorological data from the NASA data assimilation system, *J. Geophys. Res*., **109**, D18312, doi:10.1029/2004JD004554.

Lokupitiya, E., A. S. Denning, K. Paustian, I. T. Baker, K. Schaefer, S. Verma, T. Meyers, C. Bernacchi, A. Suyker, and M. Fischer, 2009. Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands. *Biogeosciences*, **6**, 969-986.

Lokupitiya, R. S., A. S. Denning, D. Zupanski, N. Parazoo, S. R. Kawa, I. Baker, and M. Zupanski, 2011. Derivation of carbon fluxes for 2003 and 2004 using a maximum likelihood ensemble filter. Submitted to *Tellus.*

Lokupitiya, E., S. Denning, K. Schaefer, D. Ricciuto, R. Anderson, M. A. Arain, I. Baker, A. G. Barr, G. Chen, J.M. Chen, P. Ciais, D.R. Cook, M. Dietze, M. El Maayar, M. Fischer, M.A. Gonzalez-Meler, R. Grant, D. Hollinger, C. Izaurralde, J.D. Jastrow, G. Katul, C. Kucharik, B. Lesht, L. Longhui, J. Liu, R. Matamala, P. Peylin, D. Price, A. Richardson, S.W. Running, A. Sahoo, M. Sprintsin, H. Tian, C. Tonitto, M. Torn, Hans Verbeeck, S.B. Verma, Y. Xue, 2011. Evaluation of model-predicted carbon and energy fluxes from cropland ecosystems. Submitted to *Biogeosciences*.

Oda, T. and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543-556, doi:10.5194/acp-11-543-2011, 2011.

Parazoo, N., A. S. Denning, R. Kawa, K. Corbin, R. Lokupitia, I. Baker, and D. Worthy, 2008. Mechanisms for synoptic transport of CO2 in the midlatitudes and tropics. *Atmos. Chem. Phys. Discussions* 2008-12197-12225.

Parazoo, N. C., A. S. Denning, J. A. Berry, A. Wolf, D. A. Randall, S. R. Kawa, O. Pauluis, and S. C. Doney, 2011. Moist synoptic transport of CO2 along the mid−latitude storm track. Geophys. Res. Let., **38**, L09804, doi:10.1029/2011GL047238.

Randerson, J. T., G. R. van der Werf, L. Giglio, G. J. Collatz, and P. S. Kasibhatla. 2007. Global Fire Emissions Database, Version 2 (GFEDv2.1). Data set. Available on-line [http://daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/849.

Schaefer, K., P. Tans, S. Denning, I. Baker, J. Berry, L. Prihodko, N. Suits, and A. Philpott, 2008. The combined Simple Biosphere/Carnegie-Ames-Stanford Approach (SiBCASA) Model, *Jour. Geophys. Res*., **113**, G03034, doi:10.1029/2007JG000603.

Schuh, A. E., A. S. Denning, K. D. Corbin, I. T. Baker, M. Uliasz, N. Parazoo, A. E. Andrews, and D. E. J. Worthy, 2010. A regional high-resolution carbon flux inversion of North America for 2004. *Biogeosciences*, **7**, 1625–1644.

Stockli, R., T. Rutishauser, D. Dragoni, P. E. Thornton, L. Lu, and A. S. Denning, 2008. Remote sensing data assimilation for a prognostic phenology model. *Jour. Geophys. Res*. **113**, G04021, doi:10.1029/2008JG000781.

Stöckli, R., T. Rutishauser, I. Baker, M. A. Liniger, and A. S. Denning (2011), A global reanalysis of vegetation phenology, *J. Geophys. Res*., **116**, G03020, doi:10.1029/2010JG001545

Takahashi, T., S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. Chavez, A. Watson, D. C. E. Bakker, U. Schuster, N. Metzl, H. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, C. Sabine, J. Olafsson, Th. S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, Richard Bellerby, A. Körtzinger, T. Steinhoff, M. Hoppema, H. J. W. de Baar, C. S. Wong, Bruno Delille and N. R. Bates (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

Zupanski, M. (2005), Maximum likelihood ensemble filter: Theoretical aspects. Mon. Wea. Rev., 133, 1710-1726.

Zupanski, D., A. S. Denning, M. Uliasz, M. Zupanski, A. E. Schuh, P. J. Rayner, W. Peters, and K. D. Corbin, 2007. Carbon flux bias estimation employing the Maximum Likelihood Ensemble Filter (MLEF). *Jour. Geophys. Res.* **112**, D17107, doi:10.1029/2006JD008371.