

## **Estimation of Regional CO<sub>2</sub> Budgets and Biomass by Fusion of LandSat, MODIS, and Atmospheric Observations**

P.I.: Scott Denning<sup>1</sup>; Co-I: Dusanka Zupanski<sup>1</sup>

Collaborators: , G. James Collatz<sup>2</sup>; Samuel Goward<sup>3</sup>; S. Randall Kawa<sup>2</sup>, Scott Doney<sup>4</sup>, Kevin Gurney<sup>5</sup>

<sup>1</sup>Department of Atmospheric Science, Colorado State University

<sup>2</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland

<sup>3</sup>University of Maryland, College Park, MD

<sup>4</sup>Woods Hole Oceanographic Institution

<sup>5</sup>Purdue University, West Lafayette, Indiana

### **Abstract**

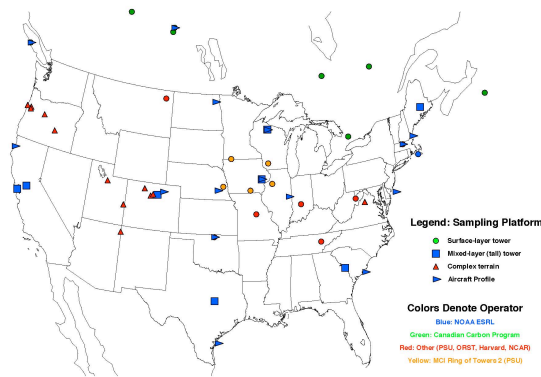
We propose an analysis of the carbon fluxes and storage by synthesizing observations of weather, surface spectra, and high-resolution land-cover, and atmospheric CO<sub>2</sub> with a suite of mechanistic models. The analysis will be performed globally, but many of the data products we propose to use are available only over the contiguous USA, so the quality of the estimation will be enhanced here and will contribute to NACP synthesis and integration. The synthesis of these data products will result in (1) global maps of time-varying sources and sinks of atmospheric CO<sub>2</sub>, wood biomass and soil carbon that are consistent with many kinds of observations; and (2) a self-consistent process-based model of these sources and sinks that can be extrapolated beyond this region of high data density. We have developed a robust data assimilation system (the Maximum Likelihood Ensemble Filter, MLEF) that can efficiently process very large volumes of data into a complex and nonlinear forward model. We have tested the system by assimilating weather and CO<sub>2</sub> data into SiB-CASA, a global model of the interactions between ecosystems and the climate. The model has been updated to include carbon allocation, storage, and biogeochemical cycling through multiple ecosystem pools. Other recent improvements include explicit treatment of crops, high-resolution fossil fuel emissions estimates, and a prognostic phenology algorithm derived by assimilating 8 years of global MODIS data. The model simulates fluxes that result from both short-term (sub-diurnal physiology to seasonal phenology) and long-term (disturbance, succession, land management) processes, predicting carbon pools as well as hourly fluxes and high-frequency variations in atmospheric CO<sub>2</sub>.

The model will be initialized using climatology and historical land-use data, then run in forward mode for the satellite era using disturbance maps derived from LandSat-derived disturbance products obtained from the North American Forest Dynamics project. Finally, we will perform ensemble data assimilation during the period of rich continuous CO<sub>2</sub> data from NACP (2007-2012), including GOSAT/Ibuki data when they become available. We will use GEOS-5 weather analyses on a 0.5° x 0.67° grid, NAFD disturbance data, GFED fire products, VULCAN fossil fuel emissions, and daily air-sea gas exchanges derived from the WHOI ocean model to predict vegetation and soil carbon pools, fluxes, and atmospheric CO<sub>2</sub> on an hourly basis. Carbon pools will then be optimized by running ensemble assimilations of data from flux towers, atmospheric CO<sub>2</sub> measurements. The final simulations will be made available through the NACP Modeling and Synthesis Thematic Data Center.

The research will address Subelement 3 (Synthesis and Integration), and is a "successor proposal" to our 2004-2006 project that intends to provide continental synthesis for NACP.

# Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

## 1. Background and Motivation



**Figure 1: Calibrated, continuous measurements of CO<sub>2</sub> over North America, 2008**

through a combination of enhanced observing systems, diagnostic and predictive models, and an ambitious effort to develop innovative model-data fusion techniques to synthesize and integrate new information (Denning et al, 2005). *To synthesize and integrate data and modeling efforts previously funded by NASA Terrestrial Ecosystem Science, we propose to use a framework for data assimilation into a process-based global model of terrestrial carbon pools and fluxes and atmospheric transport that can be used to estimate finely resolved sources and sinks (and associated uncertainty), constrained by a wide range of observations. Particular emphasis will be placed on mechanistic representation of processes related to forest disturbance and agricultural ecosystems.*

Spatial and temporal variations in the mixing ratio of atmospheric CO<sub>2</sub> are a rich source of information about the global carbon cycle, and have been analyzed by increasingly sophisticated inverse methods to infer regional sources and sinks (e.g., Gurney et al, 2002, Rödenbeck et al, 2003; Peters et al, 2005; Baker et al, 2006). Although the planned network of instrumented tall towers planned for NACP has been slow to deploy, a cooperative network of well-calibrated CO<sub>2</sub> measurements made from flux towers, telecommunication towers, and mountaintops has emerged in the past two years, with hourly data now available from nearly 50 locations in North America (Figure 1). Continuous measurements over the continent provide an opportunity to analyze CO<sub>2</sub> variations on synoptic time scales due to passing weather disturbances, opening a window into the integrated behavior of ecosystems on scales of a 10<sup>2</sup> to 10<sup>3</sup> km, which were previously unresolvable in global inverse models (Lin et al, 2006; Gerbig et al, 2006). These emerging observations (as well as “campaign-mode” measurements made from aircraft) have been used by a series of recent regional inverse studies (Gerbig et al, 2003a,b; Peylin et al, 2005; Matross et al, 2006). The NACP community is excited by these results, because of the promise of meaningful cross-comparison between process-based bottom-up extrapolations and top-down atmospheric flux estimates at regional scales that appear to be in reach at long last.

Recent advances in atmospheric inverse modeling have emphasized avoidance of predefined space-time patterns of estimated fluxes into large static regions (as used by the TransCom experiments, Gurney et al, 2002). Any errors in the assumed space-time patterns (“aggregation error”) are unavoidably aliased into errors in the magnitude of the regional flux (Kaminski et al, 2001; Engelen et al, 2002). Given sufficient data, aggregation error can be greatly reduced by solving for fluxes on the smallest possible spatial grid and at the highest possible temporal frequency, though this approach necessarily entails greatly increased computational cost relative to the coarse resolutions in space and time that have been applied in the past. Backward-in-time transport from “receptors” defined at the time and location of each observation can reduce the computational cost of high-resolution inverse calculations (e.g., Uliasz et al, 1996; Kaminski et al, 1999; Rödenbeck et al, 2003; Gerbig et al, 2003b; Peylin et al, 2005; Matross et al, 2006; Zupanski et al, 2007). In practice, the

observational constraint for such calculations is still quite weak, so that meaningful information about upstream surface fluxes is only obtained fairly close to the time and location of the measurements.

Estimation of CO<sub>2</sub> fluxes at the grid scale from atmospheric data generally requires very aggressive post-aggregation to much larger regions. Rödenbeck et al (2003) smoothed gridded flux estimates with an autocorrelation length scale of 0.2 times the radius of the Earth over land, for example, though wildly heterogeneous fluxes are known to exist over land. Michalak et al (2005) also assumed fluxes to vary smoothly over large areas, but solve for the correlation length scale as part of their inversion. Worse, temporal aggregation errors have scarcely been addressed by inversion studies to date. Rödenbeck et al (2003) aggregated CO<sub>2</sub> mixing ratios to monthly means, and estimate surface fluxes only on monthly time scales as well. Peylin et al (2005) estimated daily mean fluxes on a 50-km grid over Europe for a month, neglecting diurnal variations. Local observations show terrestrial fluxes and concentration anomalies changing sign on diurnal time scales in synchrony with systematic changes in atmospheric mixing and convection (e.g., Baldocchi et al, 2003; Bakwin et al, 1998; Gerbig et al, 2003a). Gerbig et al (2003b) and Matross et al (2006) allowed for diurnal variations in fluxes driven by photosynthetically active radiation (PAR), but aggregated parameters in their model to a handful of vegetation types, neglecting spatial heterogeneity in biogeochemistry.

Modern data assimilation techniques (e.g., Variational or ensemble methods) must be used to reduce the computational dimensions and cost of the inverse problem in a data dense world (Kalnay et al, 2003; Peters et al, 2005; Baker et al, 2006; Lokupitiya et al, 2008). Assimilation into coupled models of surface carbon exchange processes and atmospheric transport may also alleviate some of the worst of the aggregation errors that plague traditional synthesis or adjoint analyses, because temporal and spatial covariance are modeled according to process and can be constrained by other observations. Though temporal autocorrelation of surface carbon fluxes may only have a time scale of hours (due to rapidly changing radiation inputs, for example), the errors in a reasonable forward model of these fluxes may have temporal coherence on time scales of weeks or even months. This is especially important since any given region is only be “visible” to the observing network during the fraction of the time that air mass trajectories carry tracers to an observing site: a realistic forward model must essentially interpolate fluxes in that region between such times.

An alternative approach to the variational and adjoint/synthesis techniques (e.g., Peylin et al, 2005a; Rayner et al, 2005) that have been applied to coupled carbon assimilation studies to date is an emerging generation of methods called Ensemble Data Assimilation (EnsDA), that have been developed for meteorological data assimilation. A major advantage of EnsDA methods is that computation of the adjoint of the forward process model is not required (Peters et al, 2005, 2007; Zupanski et al, 2007). Estimation of parameters and uncertainty in forward coupled models of arbitrary complexity can be performed. Moreover, formal estimation of model error is possible, and may even be required for to ensure unbiased estimation (Fletcher and Zupanski, 2007). The recently implemented NOAA CarbonTracker system (Peters et al, 2007; <http://www.esrl.noaa.gov/gmd/ccgg/carbontracker>) consists of modules for the forward simulation of NPP and decomposition on land, fires, fossil fuel emissions, and air-sea gas exchange which are adjusted based on CO<sub>2</sub> anomalies observed at flask and in-situ stations.

Using data assimilation methods, it is possible to estimate the net time-mean surface carbon flux from variations in atmospheric CO<sub>2</sub> without representing all the governing processes. Two caveats must be emphasized here: (1) It is possible that overfitting of parameters in a simplistic model will reproduce observed variations in atmospheric CO<sub>2</sub> for the wrong reasons (e.g., by tuning physiological parameters such as light-response or drought stress functions) when the true source or sink results from processes that are not represented in the model (e.g., land-use change or nutrient deposition); and (2) it is crucial to be as accurate as possible with that subset of processes that control variations on the time/space scales present in the observations, especially if they covary with transport. At a minimum, coupled carbon data assimilation models relying on this separation of time scales to estimate time-mean fluxes from atmospheric CO<sub>2</sub> must represent ecophysiological processes on diurnal and synoptic time scales. Source/sink attribution to slower processes such as management, disturbance, succession, fertilization, and/or climate change also depends on credible

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

constraint of fossil fuel emissions and biomass burning, and can be improved by using remote sensing to constrain seasonal and interannual variations in ecosystem states.

Building on our successful efforts over the past two years, we propose to develop a generalized framework for ensemble data assimilation into a coupled model of the terrestrial carbon cycle and overlying dynamic atmosphere, including both “fast” ecophysiological processes that can be realistically simulated and slower carbon turnover due to disturbance and succession. The forward coupled model will be built from mature existing components, reducing development time and effort. A key innovation in the proposed data assimilation system will be an emphasis on assimilation of *state variables* (stored carbon pools) in addition to model *parameters* (e.g., allocation and turnover time). This innovation will allow us to analyze the global carbon cycle in terms of mechanistic processes rather than arbitrary covariance length scales, yet retain fidelity to atmospheric and remotely sensed data. Spatial patterns and high-frequency variations will be derived from MODIS products and analyzed weather. Slowly varying carbon pools will be predicted over time using biogeochemical principles, with forest disturbance and recovery specified from time-marched LANDSAT imagery analyzed by the North American Forest Disturbance (NAFD) project, and constrained by forest inventory analyses. The forward model will be integrated on a global 0.5° x 0.67° grid (about 50 km spacing), but forest disturbance and recovery dynamics will be based on underlying LANDSAT data at much higher resolution.

### Objectives

The central objectives of the proposed research is to *develop and evaluate a framework for synthesis of historical and current satellite imagery and high-frequency atmospheric CO<sub>2</sub> observations in support of NACP, and to apply this method to quantitative estimation of surface carbon fluxes and their uncertainty at regional scales*. In support of these overarching objectives, we identify the following specific component tasks to be performed in the course of the project:

1. Develop and evaluate an improved forward model of terrestrial carbon exchange, including prognostic phenology, agricultural ecosystems, dynamic forests, and fire;
2. Parameterize the improved biogeochemical model over North America using carbon pools estimated from LANDSAT data cubes forest inventory data by the NAFD project;
3. Analyze carbon pools and fluxes over most of North America by ensemble assimilation of CO<sub>2</sub> observations into the new model using an offline atmospheric transport; and
4. Evaluate the resulting pools, fluxes, and mixing ratios using a suite of in-situ observations.

Each of these tasks is described in more detail in the following section.

## 2. Results from Previous Research (awards NNG05GD15G and NNX08AM56G)

### 2.1. Coupled physiology & biogeochemistry (SiB-CASA)

The Simple Biosphere model (SiB) is based on a land-surface parameterization scheme originally used to compute biophysical exchanges in climate models (Sellers et al., 1986), but later adapted to include ecosystem metabolism (Sellers et al., 1996a; Denning et al., 1996a). The parameterization of photosynthetic carbon assimilation is based on enzyme kinetics originally developed by Farquhar et al. (1980), and is linked to stomatal conductance and thence to the surface energy budget and atmospheric climate (Collatz et al., 1991, 1992; Sellers et al., 1996a; Randall et al., 1996). The model has been updated to include prognostic calculation of temperature, moisture, and trace gases in the canopy air space, and the model has been evaluated against eddy covariance measurements at a number of sites (Baker et al., 2003; Hanan et al., 2004; Vidale and Stöckli, 2005). Other recent improvements include biogeochemical fractionation and recycling of stable carbon isotopes (Suits et al., 2004), improved treatment of soil hydrology and thermodynamics, and the introduction of a multilayer snow model based on the Community Land Model (Dai et al., 2003). Direct-beam and diffuse solar radiation are treated separately for calculations of photosynthesis and transpiration of sunlit and shaded

canopy fractions, using algorithms similar to those of DePury and Farquhar (1997). The model is now referred to as SiB3.

Until recently, ecosystem respiration was treated in SiB by scaling a temperature and a moisture response to achieve net carbon balance at every grid cell in one year by prescribing the size of a single pool of organic matter. This approach has recently been replaced by a scheme for allocation, transformation, and decomposition based on the Carnegie/Ames/Stanford Approach (CASA, Randerson et al., 1997). Stored photosynthate is allocated to leaves, stems, and roots in fractions that are constrained by changes in satellite vegetation index (NDVI). Carbon is tracked through biomass pools and released to the surface as dead litter, woody debris, and root litter, where it interacts with a microbial pool to produce several pools of soil organic matter and CO<sub>2</sub>. The interactive biogeochemistry module has been tested at dozens of eddy-covariance sites and found to improve simulations of the seasonal cycle of net ecosystem exchange relative to the single-pool model it replaces (Schaefer et al, 2008). Following previous work with CASA (van der Werf et al, 2005), we also plan to add a fire module to this model.

Historically, SiB has used prescribed vegetation parameters derived by remote sensing (Sellers et al., 1996b). At global scales, this approach allows realistic simulation of spatial and temporal variations in vegetation cover and state (Denning et al., 1996; Schaefer et al., 2002, 2005). At the underlying pixel scale, however, phenology products derived from satellite data must be heavily smoothed to remove dropouts and artifacts introduced by frequent cloud cover. An inevitable trade-off between cloud-induced “noise” in the leaf area and time compositing systematically stretches the seasonal cycle by choosing data late in each compositing period in spring, and early in each composite in fall. Under separate support from NASA’s Energy and Water System program, we have addressed this problem by developing and testing a prognostic phenology module for SiB. We assimilated MODIS fPAR and LAI into the prognostic phenology model to estimate its parameters (e.g., growing degree day thresholds), and produced an algorithm to predict these variations rather than prescribing them directly from imagery (Stockli et al, 2008). Finally, the revised model includes an explicit treatment of the phenology and physiology of agricultural crops, and predicts crop yields and biomass (Lokupitiya et al, 2009).

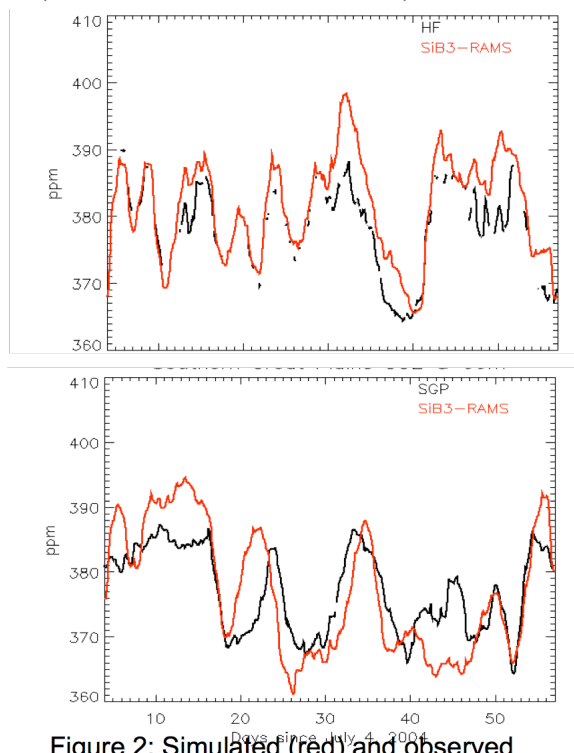


Figure 2: Simulated (red) and observed (black) CO<sub>2</sub> mixing ratios at Harvard Forest (top) and the ARM-SGP site in Oklahoma (bottom), during the summer of 2004. The diurnal cycles have been removed to highlight synoptic variations.

## 2.2 Coupled physiology, weather, and transport (SiB-RAMS)

SiB has been successfully coupled to the Regional Atmospheric Modeling System (RAMS) and used to study PBL-scale interactions among carbon fluxes, turbulence, and CO<sub>2</sub> mixing ratio (Denning et al., 2003) and regional-scale controls on CO<sub>2</sub> variations (Nicholls et al., 2004; Corbin, 2005; Wang et al, 2006; Corbin et al, 2007). Photosynthesis in SiB3-RAMS is parameterized from 1-km resolution LAI and fPAR data from NASA's MODIS satellite, and the vegetation cover is derived from 1-km MODIS land classification data (Zhao et al., 2005). The CO<sub>2</sub> concentration field at the initial time and lateral boundaries is prescribed from a global analysis using the Parameterized Chemical Transport Model (PCTM), driven from the NASA Goddard EOS Data Assimilation System on a 1° x 1.25° grid, with 20 vertical levels (Kawa et al, 2004; Parazoo, 2007).

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

The meteorological fields for RAMS are initialized and driven by the National Center for Environmental Prediction (NCEP) Mesoscale Eta-212 grid reanalysis with 40 km horizontal resolution; and lateral boundary and interior nudging is performed for horizontal wind speed, relative humidity, air temperature, and geopotential height.

The coupled model was integrated for four months during the growing season of 2004, on a 40-km grid covering most of North America. The analysis was evaluated against observations of ecosystem fluxes of heat, water, and CO<sub>2</sub> at a large number of flux towers, and the simulated CO<sub>2</sub> concentrations compared reasonably well with continuous observations at eight tower sites (Figure 2).

### 2.3 Maximum Likelihood Ensemble Filter (MLEF)

We have developed a method for regional CO<sub>2</sub> flux inversion using a Lagrangian Particle Dispersion Model (LPDM) driven by the output of SiB-RAMS. The method involves four steps: (1) forward simulation of photosynthesis, respiration, decomposition, and atmospheric transport using the coupled SiB-RAMS model; (2) calculation of a large number of backward-in-time particle trajectories from each observation point (“receptor”) in space and time; (3) integration of the particle trajectories to quantify the “influence function” of each upstream grid cell at each previous time with respect to a particular observation; and (4) an optimization scheme that adjusts the fluxes so that simulated and observed mixing ratios differ by acceptable amounts. This method was tested using data from the Ring of Towers experiment (Zupanski et al, 2007). The LPDM (Uliasz and Pielke, 1991; Uliasz, 1993, 1994; Uliasz et al., 1996) accounts for transport by resolved advection and by subgrid-scale stochastic motion (turbulence and clouds). Influence functions calculated by integrating upstream contact time with the surface quantify the partial derivative of a particular measurement with respect to all previous fluxes at all surface points in the domain (the method is nearly identical to that of Gerbig et al., 2003b). In general, influence functions are also calculated with respect to the initial distribution of CO<sub>2</sub> and the lateral boundary conditions, though with sufficient integration time the former become negligible.

We account for high-frequency time variations of photosynthesis and respiration by assuming that they are driven by well-understood and easily modeled processes (radiation, temperature, soil moisture), then solve for unknown multiplicative biases in each component flux after smoothing in space and time. This is accomplished by convolving the influence functions generated from LPDM with gridded photosynthesis (gross primary production, GPP) and ecosystem respiration (RESP) at each time step in SiB-RAMS. The net ecosystem exchange (NEE) is composed of these two component fluxes:

$$NEE(x,y,t) = RESP(x,y,t) - GPP(x,y,t) \quad (\text{eq 1})$$

where  $x$  and  $y$  represent grid coordinates and  $t$  represents time. Sub-hourly variations in the simulated component fluxes in time are primarily controlled by the weather (especially changes in radiation due to clouds and the diurnal cycle of solar forcing), whereas seasonal changes are derived from phenological calculations parameterized from satellite imagery. Fine-scale variations in space are driven by variations in vegetation cover, soil texture, and soil moisture. To estimate regional fluxes from atmospheric mixing ratios, we assume that the model of the component fluxes is biased, and that the biases are smoother in time and space than the fluxes themselves:

$$NEE(x,y,t) = \beta_{RESP}(x,y)RESP(x,y,t) - \beta_{GPP}(x,y)GPP(x,y,t) \quad (\text{eq 2})$$

A persistent bias in photosynthesis might result from underestimation of leaf area, available nitrogen, or soil moisture, whereas a persistent bias in respiration might result from overestimation of soil carbon or coarse woody debris. In any case, it is reasonable that such biases vary much more slowly than the fluxes themselves.

To estimate slowly-varying biases  $\beta_{RESP}$  and  $\beta_{GPP}$  using SiB-RAMS and LPDM, we first generate surface flux influence functions by integrating the backward-in-time particle trajectories from LPDM. Using these, we can represent the mixing ratio observed at a given station  $k$  at time  $m$  as

$$C_{k,m} = \sum_{i,j,n} \left( (\beta_{R,i,j} \text{RESP}_{i,j,n} - \beta_{A,i,j} \text{GPP}_{i,j,n}) C_{k,m,i,j,n}^* \right) \Delta t_j \Delta x \Delta y + C_{BKGD,k,m} \quad (\text{eq 3})$$

where  $i$  and  $j$  are grid indices in the zonal and meridional directions,  $n$  is the time at which GPP and Respiration occurred (not usually the time at which the resulting change in mixing ratio was measured!). The influence function  $C_{k,m,i,j,n}^*$  is then the discrete form of the partial derivative of the observed mixing ratio with respect to the NEE at grid cell  $(i,j)$  at time step  $n$ . The length scales  $Dx$  and  $Dy$  are the sizes of the grid cells in the zonal and meridional direction, and  $\Delta t_j$  is the time step over which the fluxes are applied. The term  $C_{BKGD,k,m}$  represents the contribution of “background” CO<sub>2</sub> flowing into the model domain from the larger scales. With a little algebra and a healthy dose of computer time, we obtain a simpler representation more practical suitable for optimization:

$$C_{obs} = \sum_{cell=1}^{nCell} \beta_{RESP,cell} C_{RESP,obs,cell}^* + \sum_{cell=1}^{nCell} \beta_{GPP,cell} C_{GPP,obs,cell}^* + C_{BKGD,obs} \quad (\text{eq 4})$$

where  $obs$  is an observation number (combines indices  $k$  and  $m$ ), and  $cell$  is a grid cell number (combines indices  $i$  and  $j$ ). The influence functions have been convolved with the GPP and RESP terms from the forward model and integrated over the time period over which the bias terms are assumed to apply:

$$\begin{aligned} C_{RESP,obs,cell}^* &= \Delta t_j \Delta x \Delta y \sum_n \text{RESP}_{cell,n} C_{obs,cell,n}^* \\ C_{GPP,obs,cell}^* &= -\Delta t_j \Delta x \Delta y \sum_n \text{GPP}_{cell,n} C_{obs,cell,n}^* \end{aligned} \quad (\text{eq 5})$$

We have experimented successfully with 10-day time scales for the bias terms, which allow influence functions on hourly fluxes and observations to be integrated for 240 hours. This approach has two important advantages: (1) the area and strength of upstream influence over 10 days is much greater than for a single hour, so the inverse problem of estimating the bias terms  $\beta$  is much better constrained than the estimation of the fluxes themselves; and (2) the storage of the influence functions in (eq 5) is 240 times smaller than would be required to store all the  $C_{obs,cell,n}^*$ !

We have implemented the model described above into the Maximum Likelihood Ensemble Filter (Zupanski, 2005; Fletcher and Zupanski, 2006), which is closely related to the Ensemble Kalman Filter (Peters et al, 2005). The MLEF is very flexible, allowing for nonlinear models of arbitrary complexity and for non-Gaussian errors. It has been adapted for separate estimation of model error as well as optimal control parameters. The essence of the ensemble data assimilation approach is that an ensemble of sets of systematically perturbed control parameters (the  $\beta$ 's in our case) are generated by the algorithm from an initial forward simulation and calculation of model-data mismatch. An ensemble of forward model integrations is then performed, and the optimization algorithm estimates values and uncertainties of each control parameter from the resulting dependence of model-data mismatch on parameter values, subject to specified prior values and error covariance.

A key advantage of the estimation of  $\beta(x,y)$  using the MLEF is that spatial covariance and correlation between biases in GPP and respiration can be propagated from one 10-day “assimilation cycle” to the next, so that spatial patterns in the bias emerge over time. In any given time window, the model is terribly underconstrained by observations, but the system “learns” about the model biases and their spatial structure over successive cycles as new observations are assimilated. Without spatial patterns of error covariance, inverse methods are prone to creating unrealistic flux patterns determined by the placement of the observations. Alternatively, one can assume that model biases are determined uniquely by vegetation type (Gerbig et al, 2003b, 2005), but this risks extreme aggregation error. Biases due to incorrect soil nitrogen or forest stand age, for example, are very unlikely to be constant across all pixels of a given vegetation type.

## Synthesis of Regional CO2 Budgets and Biomass

We have evaluated the ability of the MLEF to estimate biases in SiB-RAMS fluxes given influence functions generated by the LPDM using synthetic observations for the Ring of Towers experiment in the summer of 2004 (Zupanski et al, 2007). A forward simulation of a 70-day period starting on 1 June, 2004 was performed in SiB-RAMS on a domain somewhat larger than the conterminous USA on a grid of  $\Delta x=40$  km. A finer nest was run on a 1000 km x 1000 km subdomain centered on WLEF with  $\Delta x=10$  km. Influence functions were generated by running the LPDM backward in time for two-hour mean “samples” from six surface layer towers in the Ring, plus five levels on the WLEF tower (all but the 11 m level). We then sought to estimate the bias factors every 10 days (seven assimilation cycles) on a 20-km grid over a 600 x 600 km area centered on the tall tower.

We created a “true” field of the biases in SiB-RAMS simulations of GPP and ecosystem respiration ( $\beta_{RESP}$  and  $\beta_{GPP}$ ) by dividing the domain in half. On the east side, we set the mean value of both  $\beta$ 's to be 1.1, and on the western half we set them to 0.5. To make the problem more difficult, we also included random deviations in each  $\beta$  chosen from a Gaussian distribution with a mean of 0 and a standard deviation of 0.1. Because we are gluttons for punishment, we applied these deviations with different decorrelation length scales: 80 km in the southern and 160 km in the northern halves of the domain. We then used these perturbed  $\beta$ 's to generate synthetic mixing ratio data by multiplying them by the LPDM influence functions (eq 6), which were already convolved with the modeled photosynthesis and respiration from SiB-RAMS, as two-hourly averages assuming that the model bias is constant over periods of 10-days. The data were also perturbed by Gaussian noise, with a mean of 0 and a variance that depended on tower height and time of day. The error assigned to the data ranged from 1 ppm above 200 m during daytime to 45 ppm below 50 m at night. Note that this formulation only allows about three “observations” per day from the surface-layer towers under well-mixed conditions, and very strongly deweights night-time and transitional values.

As a “first guess” of the unknown distribution of model bias, we assumed a uniform field of  $\beta = 0.75$  in every grid cell. This value was assumed to be known to within 0.2 (at  $1\sigma$ ). Our initial estimate of the spatial decorrelation length-scale was 120 km. Successive cycles in the assimilation used the estimated  $\beta$ 's and covariance matrix from the previous cycle as a background field, constituting a “persistence forecast” for both the  $\beta$ 's and their covariance structure. No further smoothing was applied. After the first cycle, the spatial covariance of the errors in  $\beta$ 's was determined from the synthetic mixing ratio data. Results (Fig 3) are very encouraging. The estimated  $\beta(x,y)$  clearly distinguish the east-west structure in the “true” field, and also capture much of the random finer variations, including the smoother patterns in the south than the north. The constraint is weak over the Great Lakes, because both GPP and Resp are zero there. Overall uncertainty in the model bias was less than 5% over most of the interior of the Ring.



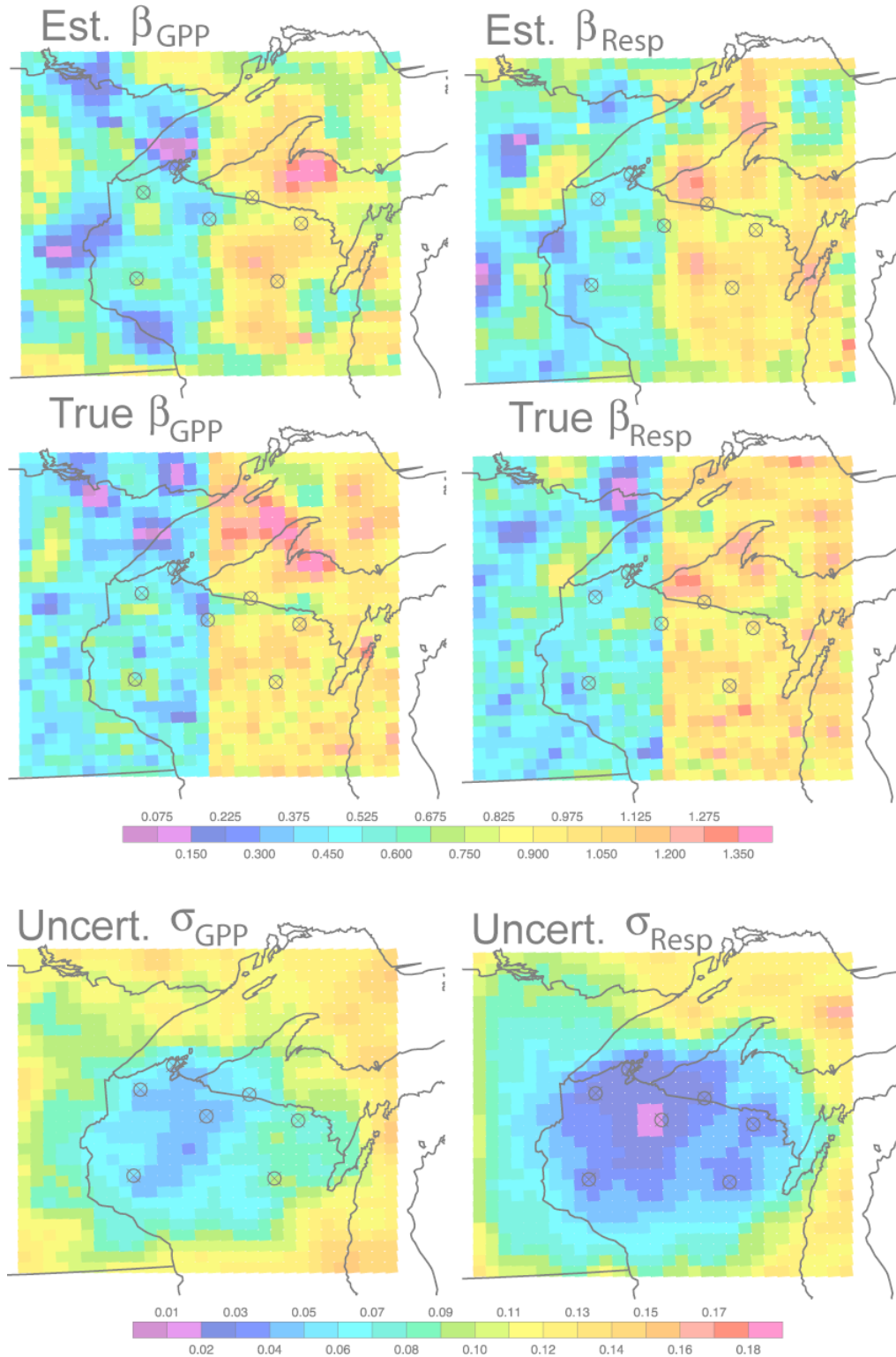


Figure 3: Assimilation of synthetic Ring  $[CO_2]$  using SiB-RAMS-LPDM-MLEF. Top panels show gridded estimates of biases in GPP ( $\beta_{GPP}$ ) and ecosystem respiration ( $\beta_{Resp}$ ) on a 20-km grid after 7 assimilation cycles of 10 days each. Small circles with X's indicate sampling sites. Prior guesses for both biases was a uniform value of 0.75. Middle panels show the prescribed “true” distribution. The lower panels show the estimated uncertainty ( $1\sigma$ ). See text for details.

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

### 3. Research Plan:

#### 3.1. Parameterization of a Biogeochemical Analysis System

A fundamental limitation of the ensemble data assimilation system described above is that we are essentially optimizing *parameters* (the gridded bias factors  $\beta$ ), but we have no mechanistic way to describe the evolution of these parameters. Because  $\beta$  is by definition an unknown quantity, we use a simple “persistence forecast” as a first guess of the value in the subsequent assimilation cycle, and use the new observations to update it. A far more powerful approach is to treat these biases as *state variables*, whose time evolution the model predicts using dynamical equations. Our development of the SiB-CASA logic has enabled this step in the continuation of this work.

We have added a biogeochemistry module to SiB3, based largely on the Carnegie-Ames-Stanford Approach (CASA, Potter et al., 1993; Randerson et al, 1997; Schaefer et al, 2007). An allocation parameterization partitions GPP into autotrophic respiration at an hourly time step and into living biomass pools (leaves, roots and stems) at a daily time step. Allocation is constrained with satellite observations of LAI and fractional woody coverage. Carbon enters non-living organic matter pools on a daily time step through the delivery of biomass to litter (leaf, root and coarse woody debris) pools. Fixed carbon is then respired back to the atmosphere and delivered to soil carbon pools controlled by pool-specific rate constants, which are scaled by temperature and moisture conditions at an hourly time step. Important parameters that control the GPP flux are the maximum biochemical capacity for CO<sub>2</sub> fixation by photosynthesis, the fraction of solar radiation absorbed by the canopy and the degree of water stress. The parameters that characterize the temperature and soil moisture response of decomposition are important determinants of the respiration fluxes. Autotrophic respiration and RH are also highly dependent on carbon pool sizes, which are state variables of the model. We have already begun optimization of parameters controlling allocation of photosynthate and turnover of the various carbon pools, based on point data from flux towers and biometric studies (Schaefer et al, 2007).

Like all biogeochemical “pool” models, SiB-CASA can be initialized by performing a long integration, but this procedure results in long-term carbon balance at every grid cell (no sources or sinks). Working with Co-Investigators and collaborators at NASA Goddard Space Flight Center (Collatz and Masek), we will use a combination of historical land-use data and new LANDSAT-derived space/time analyses of forest disturbance history from 1972-2005 initialize SiB-CASA after a long spinup to equilibrium. We will begin with detailed analyses in intensively studied regions covered by the “LANDSAT data cubes” which document land-cover change every two years at 30-m resolution over limited areas (J. Masek, personal communication). These well-studied regions also include flux towers, plot-scale biometry, manipulative experiments, and other intensive

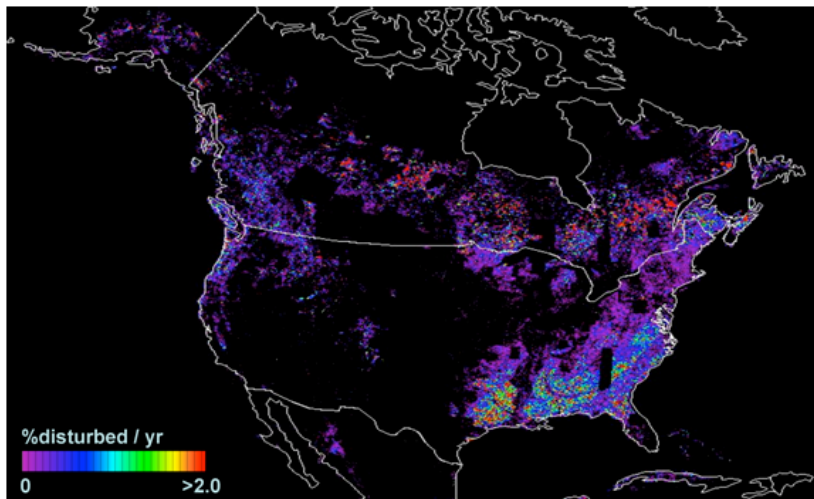


Figure 4: Initial estimate of forest disturbance rate by the LEDAPS project, 1990-2000

measurements (e.g., the WLEF tall-tower region of northern Wisconsin studied by Davis et al, the Oregon-California forest region), and model predictions of fluxes and pool sizes will be carefully compared with in-situ measurements. Gridded model estimates of biomass and carbon pools will then be compared to plot-level data derived by our collaborators (Goward and Collatz), and errors addressed.

Under separate support from DOE-NICCR, we have developed greatly improved logic for simulating the phenology, physiology, biogeochemistry, and management (tillage, harvest) of agricultural ecosystems. SiB-CASA

simulations of agroecosystem fluxes and pools will be parameterized from spatial data products generated by combining MODIS imagery with county-level agricultural statistics. The model has been evaluated using high-resolution simulations of the NACP Mid-Continent Intensive experiment in 2007-2008 (Lokupitiya et al, 2009)

Having parameterized and evaluated the biogeochemical components of SiB-CASA over the intensively-studied regions, we will then initialize the model on a 10-km grid covering most of North America using the “wall-to-wall” disturbance history products being developed by our collaborators at NASA GSFC (Masek et al, 2006; Masek and Collatz, 2006, see Fig 4). Magnitudes and spatial patterns of simulated biomass and carbon pools will be evaluated in detail against county-level statistics derived by our collaborators (Heath) from FIA data.

### 3.2. Global Atmospheric Tracer Transport (PCTM)

The Parameterized Chemistry Transport Model (PCTM) will be used for forward global simulations of CO<sub>2</sub> transport (Kawa et al., 2004; Parazoo et al, 2008). This provides a diagnostic tool for studying synoptic interactions among weather and surface CO<sub>2</sub> flux. Transport fields are provided by the NASA Goddard EOS Data Assimilation System, version 5 on a 0.5° x 0.67° grid and include 6-hourly analyzed winds and physical parameterizations determined from the National Center for Atmospheric Research Community Climate Model, Version 3 (Kiehl et al., 1998). Subgrid scale vertical processes include cumulus convection (cloud mass flux) from deep (Zhang and McFarlane, 1995) and shallow (Hack et al., 1994) parameterized convection and boundary-layer turbulence.

The model was evaluated by comparing hourly, synoptic, seasonal, and interannual variations observed at a network of 16 towers measuring calibrated continuous CO<sub>2</sub> mixing ratios on 4 continents (Parazoo et al, 2008). The model is able to realistically simulate high-frequency variations and tight spatial gradients associated with synoptic weather systems (Fig 5).

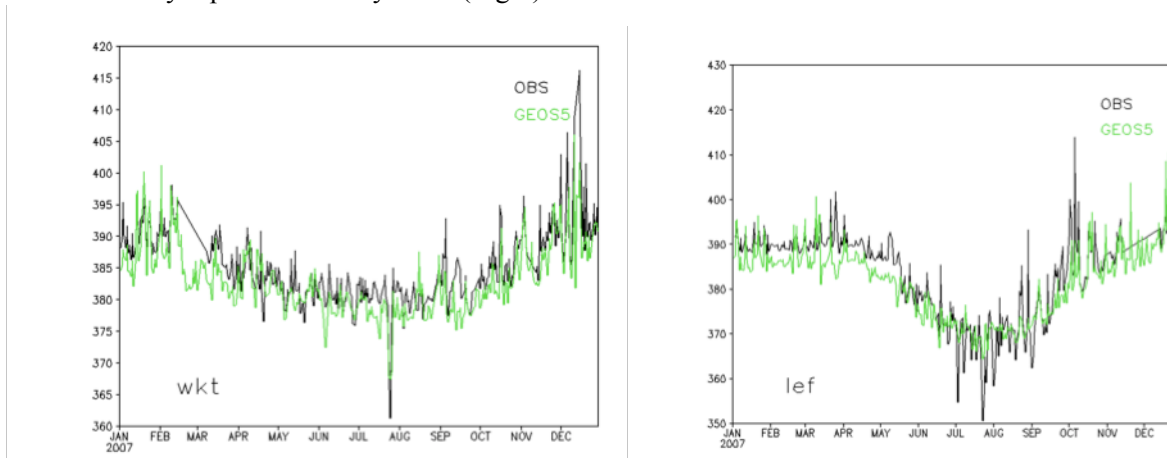


Figure 5: Simulated and observed CO<sub>2</sub> at two towers in 2007.

### 3.3. Air-sea gas exchange

Air-sea gas exchange will be obtained from a multi-decade (1979–present) hindcast experiment conducted with the Community Climate System Model (CCSM-3) ocean carbon model (Doney et al, 2009). The CCSM-3 ocean carbon model incorporates a multi-nutrient, multi-phytoplankton functional group ecosystem module coupled with a carbon, oxygen, nitrogen, phosphorus, silicon, and iron biogeochemistry module embedded in a global, three-dimensional ocean general circulation model. The model is forced with physical climate forcing from atmospheric reanalysis and satellite

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

data products and time-varying atmospheric dust deposition. Data-based skill metrics have been used to evaluate the simulated time-mean spatial patterns, seasonal cycle amplitude and phase, and subannual to interannual variability. Evaluation data include: sea surface temperature and mixed layer depth; satellite-derived surface ocean chlorophyll, primary productivity, phytoplankton growth rate and carbon biomass; large-scale climatologies of surface nutrients, pCO<sub>2</sub>, and air–sea CO<sub>2</sub> and O<sub>2</sub> flux; and time-series data from the Joint Global Ocean Flux Study (JGOFS).

### **3.4. Biomass Burning**

Emissions of CO<sub>2</sub> due to biomass burning will be specified using the Global Fire Emissions Database (GFED v2.1, Randerson et al, 2007). The 8-day emissions data set was compiled using satellite data and the Carnegie-Ames-Stanford Approach (CASA) biogeochemical model. Burned area from 2001-2004 was derived from active fire and 500-m burned area data from MODIS (Giglio et al., 2006). ATSR (Along Track Scanning Radiometer) and VIRS (Visible and Infrared Scanner) satellite data were used to extend the burned area time series back to 1997 (Arino et al., 1999; Giglio et al., 2003; Van der Werf et al., 2004). Fuel loads and net flux from terrestrial ecosystems were estimated as the balance between net primary production, heterotrophic respiration, and biomass burning, using time varying inputs of precipitation, temperature, solar radiation, and satellite-derived fractional absorbed photosynthetically active radiation. Tropical and boreal peatland emissions were also considered, using a global wetland cover map (Matthews and Fung, 1987) to modify surface and belowground fuel availability.

### **3.5. Beyond $\beta$ -- Biogeochemical Data Assimilation**

A key advantage to developing and testing the global SiB-CASA analysis is that it provides a framework for mechanistic specification of space-time patterns of fluxes in the inversion of high-frequency CO<sub>2</sub> observations. Rather than estimating gridded fluxes and applying ad-hoc spatial autocorrelation functions, or filling in large regions of fluxes like a giant game of “paint-by-biome,” we propose to estimate the state variables (biomass and carbon pools) in the model. The carefully constructed forward calculation of pools driven by MODIS land cover, analyzed weather, LANDSAT imagery, and forest inventory analysis will serve as a “strong prior” for the subsequent assimilation of CO<sub>2</sub> mixing ratio data.

High-frequency transport diagnostics (winds, turbulence, cloud mass fluxes) will be archived from the decade-long SiB-CASA-PCTM simulation on the 0.5° x 0.67° grid, and used to derive upstream influence functions for each observing station in each hour. These will then be used to optimize the woody biomass and slow soil carbon pools in SiB-CASA through the MELF assimilation system. We have found that the other pools do not need to be optimized: fast pools like leaf litter simply adjust to the fluxes, and the slowest (“armored”) pools of soil carbon contribute so little to the flux that they can be assumed to be known. The advantage of estimating pools as state variables in the model over parameter estimation is that they are controlled by the dynamical equations of the model, so that many observations over a period of years can be used to optimize them.

We have previously shown (Zupanski et al, 2007; Lokupitiya et al, 2008) that the MLEF is able to successfully recover gridded monthly “bias factors” (which we have denoted as  $\beta$ ) for both GPP and respiration, given sufficiently dense observations. By taking advantage of the NAFD disturbance maps, we will be able to initialize SiB-CASA so that instead of poorly-understood  $\beta$ , we can progressively refine spatial estimates of model state variables (biomass and soil carbon) using the atmospheric observations.

## 4. Management Plan

### 4.1. Personnel

Scott Denning is an internationally-recognized expert in carbon cycle modeling, application of remotely sensed data in carbon cycle studies, interpretation of atmospheric trace gas observations, and source/sink estimation by atmospheric inverse modeling. He will serve as the intellectual leader of the project, supervise staff, advise the graduate students, and coordinate research activity.

Dusanka Zupanski is an expert in meteorological data assimilation theory and practice, specializing in estimation of forward model error. She spent over a decade at the National Center for Environmental Prediction (NCEP), where she worked on problems related to discontinuous moist physical processes and assimilation of precipitation observations. She will lead our efforts on Ensemble Data Assimilation (EnsDA) using the coupled land-atmosphere model. She will work with other scientists and students to develop a general framework for estimation of model parameters and uncertainties using a suite of different data products.

Ian Baker is a meteorologist specializing in land surface-atmosphere interactions. He will be responsible for implementing, evaluating, and using the coupled SiB-CASA model of regional meteorology and carbon cycling. He will also assist the graduate student in the use and interpretation of the modeling system.

The graduate research assistant will work primarily on source/sink and carbon pool estimation in the coupled model, in the EnsDA framework developed by the project. They will also obtain advanced degrees and enter the scientific workforce at the end of the project as experts in coupled land-atmosphere carbon data assimilation, of whom there's currently a terrible shortage!

### 4.2. Schedule of Work

The proposed research will be performed over a period of three years. During year 1, the SiB-CASA model will be further developed and tested, initial carbon pools will be developed from the LANDSAT-derived NAFD data products over North America, and predictions evaluated against FIA data. In year 2, the transport model will be used in cycled data assimilation experiments using CO<sub>2</sub> observations at NACP towers to produce analyses over a one-year period that can be rigorously evaluated against FIA and crop production data. In year 3, we will use the mature SiB-CASA-PCTM system in the MLEF to estimate magnitudes and uncertainties of time-mean fluxes, and wood and soil carbon pools, using a broad range of real observations. Results will be disseminated through traditional means (conference presentations, journal articles, and research reports) and also by making state-of-the-art analyses available via the NACP Modeling and Synthesis Thematic Data Center through our collaborators at Oak Ridge National Laboratory.

### 4.3. Computing Requirements

We will employ Maximum Likelihood Ensemble Filter (MLEF, described in Zupanski 2005; and Zupanski and Zupanski 2006) to assimilate synthetic meteorological and carbon observations. We will apply the coupled meteorological and simple biosphere model (SiB-CASA) and PCTM within the MLEF algorithm and perform data assimilation experiments to estimate CO<sub>2</sub> fluxes using real meteorological and carbon observations. The SiB-CASA and PCTM models will be run on a 0.5° x 0.67° global grid. The computational approach includes solving numerical finite difference partial differential equations (PDE), iterative minimization of a cost function, ensemble forecasting, and evaluation of various data assimilation diagnostics (e.g., root mean square errors, degrees of freedom for signal, entropy reduction, mean, mode, covariance, and chi-square tests). The main program codes include SiB-CASA and PCTM models and MLEF data assimilation code. The main program language is FORTRAN. Unix job scripts are used to run the executables. Libraries and other software needed for the MLEF algorithm include LAPACK, BLAS, GRADS and NETCDF. All experiments will be run on multiple parallel processors. The entire MLEF algorithm is “embarrassingly” parallel, since it is based on independent runs of different ensemble members on

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

multiple processor computers. It is, therefore, highly scalable. This parallel algorithm has already been tested on Columbia computers. We do not anticipate any above average needs for technical support, since we have implemented and tested the MLEF algorithm on Columbia. A temporary storage (e.g., under nobackup2c, nobackup1, nobackup2, and nobackup3 directories) of 400 gigabytes, and a long-term storage (e.g., under home directory) of 20 gigabytes would be sufficient. However, we need a relatively high allocation of computational time on Columbia (250,000 CPU hours), due to the need for running data assimilation experiments with many ensembles of the GEOS5-PCTM model. These experiments will be more complex and computationally more expensive than in the previous years. Typically we will run experiments with 100-200 ensembles, but in some tests experiments we anticipate running up to 1000 ensemble members. We expect that a typical job will be run using 100-200 processors, and that it would last about 5-7 hours (wall-time).

### 5. References Cited

- Baldocchi D, et al, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, **82**(11), 2415–2434.
- Arino, O., J.-M. Rosaz, and P. Goloub. 1999. The ATSR World Fire Atlas. A synergy with 'Polder' aerosol products. *Earth Observation Quarterly*, 1-6.
- Baker, D. F., et al, 2006 TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO<sub>2</sub> fluxes, 1988-2003, *Global Biogeochemical Cycles*, **20**, GB1002, doi:10.1029/2004GB002439, 2006.
- Bakwin, P.S, P.P. Tans, D.F. Hurst, and C.L. Zhao, 1998: Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program. *Tellus* **50B**: 401–415.
- Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan.*, **75**, 257–288.
- Collatz, G. J., Ball, J. T., Grivet, C. and Berry, J. A., 1991. Physiological and environmental regulation of stomatal conductance, photosynthesis, and transpiration: a model that includes a laminar boundary layer. *Agric. and Forest Meteorol.*, **54**, 107-136.
- Collatz, G. J., Ribas-Carbo, M. and Berry, J. A., 1992. Coupled photosynthesis-stomatal conductance model for leaves of C<sub>4</sub> plants. *Aust. J. Plant Physiol.*, **19**, 519-538.
- Corbin, K. D., 2008. *Investigating Causes of Regional Variations in Atmospheric CO<sub>2</sub> Concentrations*. Ph.D. Dissertation, Colorado State University.
- Dai, Y., X. Zeng, R.E. Dickinson, I. Baker, G. Bonan, M. Bosilovich, S. Denning, P. Dirmeyer, P. Houser, G. Niu, K. Oleson, A. Schlosser and Z.-L. Yang, 2003: The common land model (CLM). *Bulletin of the American Meteorological Society*, **84**, 1013–1023.
- Denning, A.S., I.Y. Fung and D.A. Randall, 1995: Latitudinal gradient of atmospheric CO<sub>2</sub> due to seasonal exchange with land biota. *Nature*, **376**, 240-243.
- Denning, A.S. et al, 1996: Simulations of terrestrial carbon metabolism and atmospheric CO<sub>2</sub> in a general circulation model. Part 1: Surface carbon fluxes. *Tellus*, **48B**, 521-542.
- Denning, A.S., D.A. Randall, G.J. Collatz and P.J. Sellers, 1996: Simulations of terrestrial carbon metabolism and atmospheric CO<sub>2</sub> in a general circulation model. Part 2: Spatial and temporal variations of atmospheric CO<sub>2</sub>. *Tellus*, **48B**, 543-567.

## Scott Denning et al

- Denning, A.S. et al, 1999: Three-dimensional transport and concentration of SF<sub>2</sub>: A model intercomparison study (TransCom 2). *Tellus*, **51B**, 266-297.
- Denning, A.S., T. Takahashi and P. Friedlingstein, 1999. Can a strong atmospheric CO<sub>2</sub> rectifier effect be reconciled with a "reasonable" carbon budget? *Tellus*, **51B**, 249-253.
- Denning, A.S., et al, 2003: Simulated and observed variations in atmospheric CO<sub>2</sub> over a Wisconsin forest using a coupled Ecosystem-Atmosphere Model.. *Global Change Biology*, **9**, 1241-1250.
- Denning, A. S., et al, 2005. *Science Implementation Strategy for the North American Carbon Program*. A report of the NACP Subcommittee of the US Carbon Cycle Science Steering Group. 77 pp.
- De Pury, D. G. G. and G. D. Farquhar, Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models, *Plant, Cell, and Environment*, **20**, 537-557, 1997.
- Engelen, R.J., A.S. Denning, K.R. Gurney and G.L. Stephens. Global observations of the carbon budget: I, 2001. Expected satellite capabilities in the EOS and NPOESS eras. *Journal of Geophysical Research*, **106**, (D17), 20055-20068.
- Engelen, R. J., A. S. Denning, and K. R. Gurney, 2002: On error estimation in atmospheric CO<sub>2</sub> inversions. *Journal of Geophysical Research*, **107**, 4635, doi:10.1029/2002JD002195.
- Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Univ. Press, New York.
- Farquhar, G. D., von Caemmerer, S. and Berry, J. A., 1980. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in C<sub>3</sub> plants. *Planta*, **149**, 78-90.
- Fletcher, S. J. and M. Zupanski, A framework for data assimilation which allows for non-Gaussian errors, submitted to *Proc. Royal Soc. of London*.
- Freitas, S. R. et al, 2000. Modeling the convective transport of trace gases by deep and moist convection. *Hybrid Methods in Engineering*, **2**, 317-330.
- Gerbig C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, C. A. Grainger, 2003. Toward constraining regional-scale fluxes of CO<sub>2</sub> with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, **108**, 4756, doi:10.1029/2002JD003018.
- Gerbig C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, C. A. Grainger, 2003. Toward constraining regional-scale fluxes of CO<sub>2</sub> with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, **108**, 4757, doi:10.1029/2003JD003770.
- Gerbig, C., J. C. Lin, J. W. Munger, and S. C. Wofsy, 2006. What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes? *Atmos. Chem. Phys.*, **6**, 539–554
- Giglio, L., J. D. Kendall, and R. Mack. 2003. A multi-year active fire dataset for the tropics derived from the TRMM VIRS. *International Journal of Remote Sensing*, **24**: 4505-4525.
- Giglio, L., G. R. van der Werf, J. T. Randerson, G. J. Collatz, and P. Kasibhatla. 2006. Global estimation of burned area using MODIS active fire observations. *Atmos. Chem. Phys.*, **6**: 957-974.
- Grell, G. 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764-787.

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

- Gurney, K. R., et al. (2002), Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models, *Nature*, **415**, 626– 630.
- Gurney, K. R., et al, 2003. TransCom3 CO<sub>2</sub> inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information. *Tellus*, **55B**, 555-579.
- Gurney, K. R., et al. (2004), Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, *Global Biogeochem. Cycles.*, **18**, GB1010, doi:10.1029/2003GB002111.
- Hack, J. J. (1994), Parameterization of moist convection in the National Center for Atmospheric Research community climate model (CCM2), *J. Geophys. Res.*, **99**, 5551-5568
- Hanan, N. P., J. A. Berry, S. B. Verma, E. A. Walter-Shea, A. E. Suyker, G. G. Burba, and A. S. Denning, 2004. Model analyses of biosphere-atmosphere exchanges of CO<sub>2</sub>, water and energy in Great Plains tallgrass prairie and wheat ecosystems. *Agricultural and Forest Meteorology*, **131**, 162-179.
- Kawa, S. R., D. J. Erickson III, S. Pawson, and Z. Zhu (2004), Global CO<sub>2</sub> transport simulations using meteorological data from the NASA data assimilation system, *J. Geophys. Res.*, **109**, D18312, doi:10.1029/2004JD004554.
- IPCC (2001) *Climate Change 2001: The Scientific Basis*. A Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881 pp.
- Kalman, R., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **D82**, 35–45.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341pp, ISBN 0-521-79629-6.
- Kaminski, T., M. Heimann, and R. Giering, 1999. A coarse grid three dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO<sub>2</sub> in the 1980s, *J. Geophys. Res.*, **104**(D15), 18,555– 18,581.
- Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, **106**, 4703– 4715.
- Kaminski, T., W. Knorr, M. Heimann and P. Rayner, 2002. Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. *Global Biogeochemical Cycles*, **16**, 1066, doi:10.1029/2001GB001463.
- Kawa, S. R., D. J. Erickson III, S. Pawson, Z. Zhu, et al, Global CO<sub>2</sub> transport simulations using meteorological data from the NASA data assimilation system, *Journal of Geophysical Research-Atmospheres*, **109**(D18), 18312, doi:10.1029/2004JD004554. 2004.
- Langenfelds, R. L. et al., 2002. Interannual growth rate variations of atmospheric CO<sub>2</sub> and its  $\delta^{13}\text{C}$ , H<sub>2</sub>, CH<sub>4</sub>, and CO between 1992 and 1999 linked to biomass burning. *Global Biogeochemical Cycles*, **16**, 1048.
- Law, R. M., et al. , 1996. Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO<sub>2</sub> inversions, *Global Biogeochem. Cycles*, **10**, 783-796.
- Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting, 2002. Using high temporal frequency data for CO<sub>2</sub> inversions, *Global Biogeochem. Cycles*, **16**, 1053, doi:10.1029/2001GB001593.



- Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting, 2003. Data and modelling requirements for CO<sub>2</sub> inversions using high frequency data, *Tellus*, **55B**, 512–521, doi:10.1034/j.1600-0560.2003.0029.
- Lovenduski, N.S., N. Gruber, **S.C. Doney**, and I.D. Lima, 2007: Enhanced CO<sub>2</sub> outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, *Global Biogeochem. Cycles*, **21**, GB2026, doi:10.1029/2006GB002900.
- Lokupitiya, R. S., D. Zupanski, A. S. Denning, S. R. Kawa, K. R. Gurney, and M. Zupanski, 2008. Estimation of global CO<sub>2</sub> fluxes at regional scale using the Maximum Likelihood Ensemble Filter. *Jour. Geophys. Res.* **113**, D20110, doi:10.1029/2007JD009679.
- 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. Submitted to *Biogeosciences*.
- Lin, J. C., C. Gerbig, S. C. Wofsy, B. C. Daube, D. M. Matross, V. Y. Chow, E. Gottlieb, A. E. Andrews, M. Pathmathevan, and J. W. Munger. (2006). What have we learned from intensive atmospheric sampling field programmes of CO<sub>2</sub>? *Tellus*, **58B**, 331-343.
- Masek, J.G., E.F. Vermont, N.E. Saleous, R. Wolfe, F.G. Hall, K.F. Huemmrich, F. Gao, J. Kutler, and T.-K. Lim, 2006: A Landsat Surface Reflectance Dataset for North America, 1990-2000, *IEEE Geosci. Remote Sens. Lett.*, **3** (1), 68-72.
- Masek, J.G., and G.J. Collatz, 2006: Estimating forest carbon fluxes in a disturbed southeastern landscape: Integration of remote sensing, forest inventory, and biogeochemical modeling, *J. Geophys. Res.*, **111** (G01006), 15 pp.
- Matross, D. M., A. Andrews. M. Pathmathevan, C. Gerbig, J. C. Lin, S.C. Wofsy, B. C. Daube, E. W. Gottlieb, V. Y. Chow, J. T. Lee, C. Zhao, P. S. Bakwin, J. W. Munger, and D. Y. Hollinger (2006). Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus*, **58B**, 344-358.
- Michalak, A. M., L. Bruhwiler, P. P. Tans, A geostatistical approach to surface flux estimation of atmospheric trace gases, *Journal of Geophysical Research-Atmospheres*, **109**, D14109, doi:10.1029/2003JD004422, 2004.
- Michalak, A.M., A. Hirsch, L. Bruhwiler, K.R. Gurney, W. Peters, and P.P. Tans (2005) "Maximum Likelihood Estimation of Covariance Parameters for Bayesian Atmospheric Trace Gas Surface Flux Inversions," *Journal of Geophysical Research*. **110**, D24107, doi:10.1029/2004JD005970.
- Nicholls, M.E., A.S. Denning, L. Prihodko, P.-L. Vidale, K. Davis, P. Bakwin, 2004: A multiple-scale simulation of variations in atmospheric carbon dioxide using a coupled biosphere-atmospheric model. *Journal of Geophysical Research*, **109**, D18117. Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P. P. Tans, An ensemble data assimilation system to estimate CO<sub>2</sub> surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, **110**, D24304, doi:10.1029/2005JD006157, 2005.
- Parazoo, N., A. S. Denning, R. Kawa, K. Corbin, R. Lokupitia, I. Baker, and D. Worthy, 2008. Mechanisms for synoptic transport of CO<sub>2</sub> in the midlatitudes and tropics. *Atmos. Chem. Phys. Discussions* **2008**-12197-12225.
- Prihodko, L., A.S. Denning, N.P. Hanan, I. Baker, K. Davis, 2008. Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. *Agric. Forest Meteorol.*, **148**, 268-287,

## Synthesis of Regional CO<sub>2</sub> Budgets and Biomass

doi:10.1016/j.agrformet.2007.08.006.

- Peylin, P., D. Baker, J. Sarmiento, P. Ciais, and P. Bousquet (2002), Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO<sub>2</sub> data, *J. Geophys. Res.*, **107**, 4385, doi:10.1029/2001JD000857.
- Peylin, P., P. Bousquet, C. Le Quéré, S. Sitch, P. Friedlingstein, G. McKinley, N. Gruber, P. Rayner, and P. Ciais, Multiple constraints on regional CO<sub>2</sub> flux variations over land and oceans, *Global Biogeochem. Cycles*, **19**, GB1011, doi:10.1029/2003GB002214, 2005.
- Peylin, P., P. J. Rayner, P. Bousquet, C. Carouge, F. Hourdin, P. Heinrich, P. Ciais, and AEROCARB contributors, 2005. Daily CO<sub>2</sub> flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology. *Atmos. Chem. Phys.*, **5**, 3173–3186.
- Potter CS, Randerson JT, Field CB, et al., 1993. Terrestrial ecosystem production - a process model-based on global satellite and surface data, *Global Biogeochem Cycles* **7**, 811-841.
- Rayner, P. J., I. G. Enting, R. J. Francey, and R. Langenfelds, 1999. Reconstructing the recent carbon cycle from atmospheric CO<sub>2</sub>,  $\delta^{13}\text{C}$ , and O<sub>2</sub>/N<sub>2</sub> observations. *Tellus*, **51B**, 213-232.
- Rayner, P.J. and D.M. O'Brien, 2001. The utility of remotely sensed CO<sub>2</sub> concentration data in surface source inversions, *Geophys. Res. Lett.*, **28**, 175-178.
- Rayner P. J. et al, 2002. Observations of the carbon budget, 3, Initial assessment of the impact of satellite orbit, scan geometry, and cloud on measuring CO<sub>2</sub> from space, *J. Geophys. Res.*, **107**, 4557, doi:10.1029/2001JD000618.
- Rayner, P.J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann, Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochemical Cycles*, **19**(2), 2005.
- Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann, 2003. Time dependent atmospheric CO<sub>2</sub> inversions based on interannually varying tracer transport, *Tellus*, **B55**, 488–497.
- Schaefer, K., A.S. Denning, N. Suits, Joerg Kaduk, I. Baker, S. Los, and L. Prihodko, 2002: The effect of climate on inter-annual variability of terrestrial CO<sub>2</sub> fluxes. *Global Biogeochemical Cycles*, **16**, 1102, doi:10.1029/2002GB001928.
- Schaefer, K., G. J. Collatz, P. Tans, A. S. Denning, I. Baker, J. Berry, L. Prihodko, N. Suits, A. Philpott, The combined Simple Biosphere/Carnegie-Ames-Stanford Approach (SiB/CASA) model, submitted to *Glob. Biogeochem. Cycles*.
- Scholze, M., Rayner, P., Knorr, W., Kaminski, T., and Giering, R., 2002. A prototype carbon cycle data assimilation system (CCDAS): Inferring interannual variations of vegetation-atmosphere CO<sub>2</sub> fluxes, *Eos Trans. CG62A-05*.
- Sellers, P.J. et al., 1996a: A Revised Land-Surface Parameterization (SiB2) for Atmospheric GCMs. Part 1: Model formulation. *J. Clim.*, **9**, 676-705.
- Sellers, P. J. et al., 1996b: A Revised Land-Surface Parameterization (SiB2) for Atmospheric GCMs. Part 2: The generation of global fields of terrestrial biophysical parameters from satellite data. *J. Clim.*, **9**, 706-737.
- Subcommittee on Global Change Research, 2003. *Strategic Plan for the U.S. Climate Change Science*

## Scott Denning et al

- Program*. A Report by the Climate Change Science Program, Washington, DC: US Global Change Research Program. Available on-line <http://www.climatechange.gov/Library/stratplan2003/final/default.htm>
- Suits, N.S., A.S. Denning, J.A. Berry, C.J. Still, J.Kaduk and J.B. Miller, Simulation of carbon isotope discrimination of the terrestrial biosphere, *Global Biogeochemical Cycles*, **19**, GB1017, doi:10.1029/2003GB002141, 2005.
- Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Uliasz, M., 1993: The atmospheric mesoscale dispersion modeling system (MDMS). *J. Appl. Meteor.*, **32**, 139-149.
- Uliasz, M., 1994: Lagrangian particle modeling in mesoscale applications. *Environmental Modelling II*, ed. P. Zannetti, Computational Mechanics Publications, 71-102.
- Uliasz, M. and R.A. Pielke, 1991: Application of the receptor oriented approach in mesoscale dispersion modeling. *Air Pollution Modeling and Its Application VIII*, eds. H. van Dop and D. G. Steyn, Plenum Press, New York, 399-408.
- Uliasz, M., R.A. Stocker, and R.A. Pielke, 1996: Regional modeling of air pollution transport in the southwestern United States. *Environmental Modelling III*, ed. P. Zannetti, Computational Mechanics Publications, 145-182.
- van der Werf GR, et al, 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, J. Collatz, P.S. Kaskhatla, A.F. Arellano Jr. (2006), Interannual variability of global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys. Discuss.*, **6**, 3175-3226.
- Vidale, P.-L. and R. Stöckli, Prognostic canopy air space solutions for land surface exchanges. *Theor. And Appl. Climatol.*, **80**, 245-257, doi:10.1007/s00704-004-0103-2, 2005.
- Walko, R.L., C.J. Tremback, R.A. Pielke, and W.R. Cotton, 1995: An interactive nesting algorithm for stretched grids and variable nesting ratios. *J. Appl. Meteor.*, **34**, 994-999.
- Wang, Y.-P., and D. J. Barrett (2003), Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach: I. Using remotely sensed data and ecological observations of net primary production, *Tellus*, **55B**, 270 – 289, doi:10.1034/j.1600-0560.2003.00031.
- Wang, Y.-P., and J. L. McGregor (2003), Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach: II. The atmospheric constraint, *Tellus*, **55B**, 290– 304, doi:10.1034/j.1600-0560.2003.00030.
- Wofsy, S.C. and R.C. Harriss (2002) *The North American Carbon Program (NACP)*, Report of the NACP Committee of the U.S. Carbon Cycle Science Program. Washington, DC: US Global Change Research Program. Available on-line <http://www.carboncyclescience.gov/nacp/>
- Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274–2292.
- Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.

## Synthesis of Regional CO2 Budgets and Biomass

Zupanski, D. and M. Zupanski, 2006: Model error estimation employing ensemble data assimilation approach. *Mon. Wea. Rev.*, **134**, 1337-1354.

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.