2. Abstract

Data fusion to determine North American sources and sinks of carbon dioxide at high spatial and temporal resolution from 2004 to 2008

Co-Principal Investigators:
Kenneth J. Davis, Klaus Keller, Natasha R. Miles and Scott J. Richardson,
The Pennsylvania State University
A. Scott Denning, Colorado State University

There is strong evidence that North America terrestrial ecosystems are currently a substantial sink of carbon dioxide. The magnitude of the sink has a large range of uncertainty, we have a limited understanding of how it has varied over time, and the processes responsible for this sink are not entirely clear. Our limited understanding is linked to methodological limits, as well as limited continental data. Quantifying spatial patterns and temporal variability of carbon dioxide sources and sinks at continental to regional scales remains a challenging problem.

In response to this challenge a rapid expansion of the N. American carbon cycle observational network is underway. This expansion includes a network (AmeriFlux) of continuous, eddy-covariance based CO_2 flux measurements and a network of continuous, continental CO_2 mixing ratio observations of comparable precision and accuracy to the marine flask network. Inverse studies of the N. American carbon budget have only begun to utilize these emerging data sources directly (i.e. tower fluxes and continuous continental mixing ratio observations), and how to best utilize these data together is a topic of great uncertainty and intensive research. This is the focus area of our proposal.

We propose to continue a program of research that will turn the emerging wealth of data in N. America to our advantage. This will be accomplished by a continued collaboration between research groups at the forefronts of terrestrial boundary layer CO_2 flux and mixing ratio observations, and high resolution, land-atmosphere carbon cycle modeling. This collaboration has resulted in substantial progress towards fusion of flux and mixing ratio observations in a coupled land-atmosphere data assimilation framework. This project will further develop methods for fusion of CO_2 flux and mixing ratio observations via inverse modeling incorporating the N. American CO_2 mixing ratio observational network, forwards modeling built upon the N. American flux network, and cross-evaluation of these two approaches. Further, we will apply the methods already developed via this collaborative effort to examine interannual variability of N. American carbon fluxes from 2004 to 2008.

The research will address the following hypotheses: 1) Flux and mixing ratio observations can be merged into a consistent analysis at synoptic, seasonal, and interannual time scales; 2) The N. American CO_2 budget will be well constrained by our data analysis system; 3) The 2004-2008 record of N. American net annual terrestrial CO_2 fluxes will show a persistent net sink of carbon of location and magnitude consistent with previous estimates based on ecological inventory methods, and; 4) The same flux record will yield detectable, spatially-resolved, climate-driven interannual variability. Expected products include: 1) a growing database of flux-tower based, continuous CO_2 mixing ratio observations suitable for application to continental inversions; 2) a comprehensive analysis system for estimation of monthly CO_2 exchange across N. America at high spatial resolution; 3) significant reduction in the uncertainty in the annual net N. American CO_2 flux and its interannual variations, and; 4) spatially and temporally resolved terrestrial CO_2 fluxes and uncertainty estimates for 2004 through 2008 encompassing all of N. America. Ultimately, the results will support the development of dynamic predictions of the future carbon cycle by providing a regionally and temporally resolved multi-year record of whole continent terrestrial carbon fluxes needed to evaluate continental-scale models.

Total proposed cost: $XXX (Combined Penn State and Colorado State effort)

Budget period: 1 May, 2007 – 30 April, 2010

3. Results from Prior Research

This narrative cover progress resulting from NOAA GCC as well as relevant DOE TCP and NASA funds. We have developed and tested a LI-COR 820 based instrument for measurement of continuous, well-calibrated CO_2 mixing ratio at flux towers (Miles *et al.*, in prep). We have shown that these measurements are accurate to approximately 0.2 ppm relative to NOAA tall tower measurements. These or similar measurements are now, or soon will be made, at 13 U.S. and 3 Canadian flux towers.

We have developed and tested a method for extrapolating these surface-layer measurements of CO_2 at flux towers to atmospheric mixed-layer values under convective (daytime) conditions, creating inexpensive "virtual tall towers (VTT)." These VTT estimates have been compared to six years of actual vertical differences measured at the WLEF tall tower. We find that hourly daytime mixed-layer mixing ratios can be estimated from surface layer values and measured fluxes to within 0.5 ppm in winter, within 0.2 ppm in summer, and within 0.05 ppm in fall and spring (Butler *et al.*, in prep). Accurate extrapolation of surface-layer data to the mixed-layer allows Ameriflux towers to contribute to regional flux estimation by inversion of large-scale transport models which cannot resolve surface-layer gradients.

We have developed several different methods for estimation of continental carbon budgets from CO_2 mixing ratio observations which combine traditional weekly flask sampling with continuous in-situ measurements. This is very challenging because of the vastly greater data volume with hourly compared to weekly observations. Older methods have estimated monthly fluxes for large regions, but this leads to unacceptable bias due to errors in the assumed spatial patterns of fluxes within regions. Finer resolution is possible using mesoscale models, but variations of CO_2 at the lateral boundary conditions is required in this case. Our strategy has been to use a global model to perform relatively coarse estimation of monthly mean fluxes, and then to use the resulting optimized 4-dimensional CO_2 field as a "first guess" for lateral boundary conditions for much higher resolution inversions using a mesoscale model.

Global inversions and transport modeling have been performed with additional support from NASA using the Parameterized Chemical Transport Model (PCTM), which is driven by analyzed meteorology produced by the NASA Goddard Modeling and Assimilation Office. We are using this model to separately estimate monthly photosynthesis and respiration for 47 regions, with 10 in North America.

At the regional scale, we have developed a method to perform flux estimation on a 100 km x 100 km grid over North America using the CSU Regional Atmospheric Modeling System (RAMS) and a backward-in-time Lagrangian Particle Dispersion Model (LPDM). RAMS transport fields are archived and used by LPDM to calculate influence functions, (partial derivative of observed CO_2 variations with respect to upstream fluxes at previous times). With a continental network of 10-20 towers making hourly measurements, it is not possible to estimate fluxes every hour for every 100 km grid cell. We aggregate fluxes for 10 days at a time using the Simple Biosphere (SiB) model coupled to RAMS, which estimates photosynthesis (GPP) and respiration every 5 minutes from physiological principles and satellite imagery. We have evaluated SiB-RAMS by comparing simulated fluxes to eddy covariance measurements. We convolved the LPDM-derived influence functions separately with simulated GPP and respiration in SiB-RAMS to produce maps of the influence of each component flux at every grid cell over 10 days on the observed mixing ratio at each tower in each hour. The inverse problem was then formulated as an estimation of multiplicative model bias in GPP and respiration in SiB-RAMS for each grid cell. Optimal estimates of these biases were applied to the simulated gridded fluxes at each time step to produce time-varying maps of GPP and respiration on the 100-km grid which are consistent with the mixing ratio variations.

We found that uncertainty in GPP and respiration was substantially reduced only in a very limited region (a few hundred km radius) around each tower unless spatial error covariance structures were introduced into the optimization. We have applied a very flexible procedure based on the Maximum Likelihood Ensemble Filter (MLEF) to perform the optimization of model bias. Unlike previous studies, we allowed for generalized error covariance and did not specify an exponential decay of spatial autocorrelation with distance. We found that with sufficiently dense observing networks (e.g., the DOE-supported Ring of Towers in 2004), the method could recover complicated spatial structures in model bias quite well. On the other hand, we found that without allowing for spatially correlated model bias the current observing network at the continental scale is insufficiently dense to constrain spatial structures over many areas.

We have used observed fluxes to study the impact of uncertain model parameters in SiB on errors in simulated fluxes (Prihodko et al, in press), and showed that model skill at synoptic to seasonal time scales was often controlled by a handful of parameters. Ricciuto et al (in press) confirmed that a model with a small number of parameters could simulate daily, synoptic and seasonal flux variability well, but Ricciuto (2006) showed that even a tuned ecosystem model had limited skill in predicting interannual variability of net ecosystem-atmosphere exchange (NEE) of CO_2 across 5 eastern U.S. temperate forest AmeriFlux sites. This suggests that changes in model structure, rather than simple parameter tuning may be required to capture interannual variability. Assimilation of multi-year records from the flux towers yielded good convergence of the parameters governing photosynthesis and forest phenology, and the parameter values were similar across these sites. Convergence of parameter values governing heterotrophic respiration, however, was weak and relatively inconsistent.

We showed that synoptic to seasonal variations were coherent across a number of towers, but that mean annual fluxes were surprisingly heterogeneous, even over a small area. Different processes control variations at different time scales. Butler et al (in prep) show that spatially coherent responses to climate anomalies can influence timing of seasonal fluxes across a large region, producing widespread anomalies in CO_2 mixing ratio that should be interpretable via inverse modeling.

With support from NASA, we completely replaced the respiration logic in SiB with a set of prognostic equations for allocation and transfer of photosynthate through a series of biogeochemical pools based on the CASA model. The SiB-CASA model can now represent the effects of disturbance and management via the storage pools, and has been evaluated against a network of flux towers (Schaefer et al, submitted).

We have studied the nature of the very strong synoptic variability in CO_2 mixing ratios at continental sites using observations at six towers, the global PCTM and the coupled SiB-RAMS models. We found that variations are predominantly driven by horizontal advection rather than changes in vertical mixing, and that they can be predicted reasonably well by the models. This is encouraging for the feasibility of regional flux inversion using these models.

With NOAA ESRL support, we have developed a system for deriving influence functions for regional inverse modeling by driving the LPDM with hourly 13-km meteorological analyses produced for the Rapid Update Cycle (RUC) model.

4. Statement of Work

4.1 Introduction

The fate of anthropogenic CO_2 introduced into the atmosphere by the combustion of fossil fuels is one of the leading sources of uncertainty in projections of future climate. Coupled carbon-climate models simulate positive feedback (warming promotes additional CO_2 release to the atmosphere), but a recent comparison of 11 such models found a range of nearly 200 ppm in CO_2 and 1.5 K of warming in 2100 (Friedlingstein et al, 2006). Research leading to improved quantification and understanding of carbon sources and sinks has therefore been identified as a major priority for the US Carbon Cycle Science Program, with special focus on North America in the near term. The North American Carbon Program (NACP, Wofsy and Harris, 2002; Denning et al, 2005) involves process studies, an expanded flux measurement network, remote sensing and modeling, and inversions using new atmospheric mixing ratio observations. Cross-evaluation of models and data sources and hypothesis testing at a variety of spatial and temporal scales is envisioned within a new framework of model-data fusion.

Direct measurements of carbon exchange between the atmosphere and terrestrial ecosystems by the eddy-covariance method has been undertaken at an increasing number of sites in North America and around the world (Baldocchi et al, 2001). These measurements provide information about CO_2 fluxes and their responses to climate variations on hourly to decadal time scales, but the areas represented by the flux measurements are very small (order 1 km^2). The tower flux measurements are excellent, however, for observing temporal variability in net ecosystem-atmosphere exchange (NEE) of CO_2 representative of particular ecosystems, and for gaining local mechanistic understanding of fluxes. Flux data are being used increasingly in formal data assimilation procedures that merge flux data with terrestrial carbon cycle models (e.g. Sachs *et al.*, 2005) to estimate model parameters. This approach opens opportunities for more carefully considered extrapolation of flux measurements to larger spatial scales. It is highly unlikely, however, that the flux tower data alone, given its limited spatial sampling and the heterogeneity of ecosystem processes, can provide sound estimates of continental-scale NEE of CO_2.

At continental scales, net continental carbon exchange is fairly well quantified through inversion of atmospheric CO_2 mixing ratio measurements using tracer transport models (e.g. Baker *et al*, 2005). These methods complement eddy-covariance measurements, having the advantage of being inherently representative of the largest spatial scales with the disadvantage that spatial variations and relationships to local processes are not resolved. Successful application of atmospheric mass-balance constraints to determine regional carbon exchange across North America will require a tremendous improvement in the spatial resolution of flux estimates from inverse modeling. Most previous estimates are based on a global network of measurements conducted on weekly air samples collected in flasks, mostly from remote marine locations to maximize representativeness. Increased continental CO_2 data density and improved inversion methods are needed to achieve the desired spatiotemporal resolution.

Synthesis inversion of atmospheric observations involves forward simulation of tracer pulses from regions with prescribed patterns of flux variations in space and time (e.g., Gurney et al, 2002; Baker et al, 2005). Prescribing spatial patterns allows other forms of information to be brought to bear on the results of the inverse calculation (e.g., we expect no carbon exchange with the Greenland ice sheet). If incorrect patterns of flux variations are prescribed as *hard constraints* (not adjustable by the optimization procedure), errors in subregional variations are inevitably aliased into biases in the estimated fluxes in the regional and time mean. This “aggregation error” can be reduced by solving for fluxes on the smallest possible spatial grid and at the highest possible temporal frequency (e.g. Kaminski et al., 2001; Engelen et al., 2002), though at greatly increased computational cost relative to the coarse inversions that have been applied in the past. Backward-in-time transport from “receptors” defined at the time and location of each observation reduces the computational cost when the number of observations is smaller than the number of potential sources and sinks (e.g., Uliasz et al., 1996; Rödenbeck et al., 2003; Gerbig et al., 2003b; Peylin et al, 2005).

In practice, the observational constraint for inversions for fluxes at the grid resolution of the transport models is still quite weak, and relies on aggressive assumptions about spatial covariance. Rödenbeck et al

(2003) assumed a correlation length scale of 0.2 times the radius of the Earth, whereas Michalak et al (2004) used a geostatistical approach to solve for length scales of spatial patterns in the fluxes. Peters et al (2005) used an ensemble Kalman filter with covariance smoothing over 900 km on land and 2000 km over oceans. Gerbig et al (2003b) used aircraft data to solve for empirical parameters in a statistical model of fluxes, assuming these were representative over very broad types of vegetation ("forest" vs "crops").

The realism of the assumptions made about spatial covariance of fluxes made in these approaches are difficult to evaluate. Synoptic or seasonal variations are expected to be fairly smooth because they are driven by weather and climate, but annual mean fluxes may result from heterogeneous disturbance history or land management on much finer spatial scales. Chevallier et al (2006) analyzed spatial patterns in the error of a model of daily net carbon flux (ORCHIDEE) relative to flux tower observations and found no meaningful patterns. Temporal aggregation errors have scarcely been addressed by inversion studies to date. Rödenbeck *et al*. (2003) aggregated CO_2 mixing ratios to monthly means, and estimate surface fluxes only on monthly time scales as well. This aggressive temporal truncation is necessary for computational efficiency, but is justified only if covariance among transport, fluxes, and mixing ratio is negligible (Denning et al., 1995, 1996b, 1999). Local observations contradict this assumption, with terrestrial fluxes and concentration anomalies changing sign on diurnal, synoptic and seasonal time scales in synchrony with systematic changes in atmospheric mixing and convection (Davis et al., 2003; Gerbig et al., 2003a; Hurwitz et al., 2004; Yi et al., 2004; Bakwin *et al.*, 2004; Helliker *et al.,* 2004).

Continuous in-situ measurements of CO_2 mixing ratio over the continents offer the possibility of dramatically strengthening the observational constraint on regional fluxes relative to weekly flask sampling (Law et al, 2003; Peylin et al, 2005). Variations of CO_2 on daily to synoptic time scales over continental areas are an order of magnitude stronger than seasonal and interannual changes, and are primarily driven by changes in airmass trajectories (Hurwitz et al, 2004; Wang et al, 2006). If these huge variations can be used as a "signal" (rather than "noise") for regional flux estimation, the inverse problem will be much better constrained. Using high-frequency continental CO_2 data requires accurate specification of meteorological transport, including PBL and cloud venting processes and synoptic-scale features such as fronts. This is difficult to achieve using global models, but is possible using a nested mesoscale model such as the Regional Atmospheric Modeling System (RAMS, Denning et al, 2003; Nicholls et al, 2004; Wang et al, 2006) or high-resolution mesoscale analyses such as those produced by the NOAA Rapid Update Cycle system (RUC, Benjamin et al, 2004a,b). The density of continental CO_2 observations has dramatically increased in recent years. These data are only beginning to be used with the atmospheric synthesis inversion approaches described above. Towards this end we have developed, as noted in section 3, methods for estimation of seasonal to interannual variations in net CO_2 flux at high resolution over N. America by applying synthesis inversion of global data and embedding a nested mesoscale inverse calculation using newly available continuous observations of the mixing ratio of CO_2. We have found that even hourly observations at dozens of locations are insufficient for estimation of hourly fluxes, and that a process-based model will be required to calculate fine-scale variations in space and time that are driven by radiation, weather, and differences in vegetation. Even using a mechanistic prior model of hourly flux variations, the inverse problem is very sensitive to assumptions about space-time error covariance in space and time, and assuming errors are uncorrelated precludes a useful result.

Flux tower data have the potential to address this uncertainty about space-time covariance in terrestrial fluxes. We have, therefore, developed methods for assimilating flux tower data into terrestrial carbon cycle models, similar to the pioneering work of Braswell *et al.*, (2005). These efforts have yielded skill in simulating daily to seasonal terrestrial CO_2 fluxes, and have suggested that tower flux data may be an effective means of characterizing ecosystem functional parameters, particularly photosynthetic processes, across broad plant functional types (Ricciuto, 2006; Gerbig et al., 2003a,b). The analyses of Butler et al (in preparation) on the spring 1998 climate anomaly in North America, and of Ciais et al., (2005) on the summer 2003 climate anomaly in Europe, also suggest that at least for strong, seasonal climate anomalies, significant spatial coherence exists in terrestrial flux fields. Wang et al., (2006) simulates similar spatial coherence of surface fluxes for a strong syntopic event, and is able to show good consistency with observed atmospheric mixing ratios.

In sum, we now possess the tools needed to estimate spatially and temporally resolved terrestrial carbon fluxes over a multi-year time frame, but remain uncertain as to whether or not the data available will provide sufficient constraint, and uncertain as to which assumptions and methods will yield the most valid results. Can flux tower data, extrapolated via a terrestrial carbon cycle model and spatially explicit environmental forcing, provide a valid bottom-up constraint at continental scales for at least a subset of the temporal domain or ecophysiological parameter space of interest? Does coherence in terrestrial ecosystem carbon exchange justify smoothing the information derived from atmospheric CO_2 mixing ratio observations over space and time so as to complement the flux tower data? Does the information content of these two CO_2 measurement networks (flux and mixing ratio) overlap at intermediate spatiotemporal scales and provide complementary information where they do not overlap (diurnal cycles for flux measurements, annual continental budget for atmospheric CO_2), or are there fundamental spatiotemporal gaps in these tools that leave gaps in quantification of the N. American carbon cycle?

We propose a three-year investigation to explore these issues. We will analyze synoptic, seasonal, and interannual variations in the N. American carbon cycle, merging information from the flux tower network and continuous observations of CO_2 mixing ratio from the growing network of calibrated in-situ analyzers. We will assimilate these observations via our analytical system capable of estimating N. American CO_2 fluxes at synoptic temporal and regional spatial resolution. We will apply this analytical system to a five-year time series of N. American CO_2 flux and mixing ratio observations (2004-2008), estimating spatially and temporally resolved continental terrestrial fluxes. In so doing we will assess the uncertainty of these flux estimates and evaluate the validity of the methods employed. The ultimate goal is to provide a key analytical component of a continental observation network that will both monitor terrestrial contributions to the global carbon cycle, and provide flux estimates that can be used to evaluate our ability to predict responses of the terrestrial carbon cycle to future climatic change.

In particular, this project will add to the observational constraint by maintaining calibrated CO_2 measurement systems at five flux towers, and assisting in data interpretation, intercalibration and management at 8-11 additional flux towers with well-calibrated CO_2 sensors. We will make the observations and the quality assurance data available to other investigators through a publicly-accessible web site. We will use a new ecosystem model to predict finely-resolved space-time variations of CO_2 fluxes, assimilating measured net ecosystem exchange measurements from dozens of flux towers to optimize these predictions on a fine grid over N. America. We will then use high-resolution satellite imagery and weather analyses to predict surface flux variations every hour over a five-year period (2004-2008), and propagate these fluxes through an atmospheric transport model that predicts variations in atmospheric CO_2 mixing ratio at each continuous CO_2 observing site. Finally, we will use NOAA's finely-resolved hourly atmospheric transport data to correct model biases and estimate gridded carbon fluxes and storage that are consistent with both flux tower and atmospheric mixing ratio measurements.

The group assembled is highly qualified to perform the proposed work. The team includes strong expertise in CO_2 flux and mixing ratio observations, micrometeorology, atmospheric transport and terrestrial carbon cycle modeling, and data assimilation methods. The study is highly leveraged, building upon considerable recent advances in both methodology and observations. The proposed three years of research, while ambitious, has the potential to make significant progress towards high-level synthesis of the North American carbon cycle. The team has a strong track record of collaboration, both internally and with colleagues in the domestic and international research community. The data, methods and results will be widely disseminated. The project is targeted to squarely address NOAA GCC's research agenda.

4.2 Hypotheses and Objectives

4.2.1 Hypotheses

1. Flux and mixing ratio observations can be merged into a consistent analysis at synoptic, seasonal, and interannual time scales. Variability in fluxes at these scales will be driven by weather and climate, coherent over entire biomes, and reflected in both flux and mixing ratio measurements.

2. The N. American CO_2 budget will be well constrained by our data analysis system: Annual mean fluxes, local in character and difficult to map with flux towers, will be determined primarily by

atmospheric mixing ratio data. The flux tower record will provide constraint at temporal scales that are too short for the atmospheric inversion to be practical.

3. The 2004-2008 record of net annual terrestrial fluxes will show a persistent net sink of carbon whose location and magnitude is consistent with previous estimates based on ecological inventories.

4. The 2004-2008 record of N. American net CO_2 fluxes will yield detectable, spatially-resolved, climate-driven interannual variability, suggesting that mechanistic studies of this variability at flux towers will be describe processes and phenomena much more extensive than the tower footprints.

4.2.2 Objectives

We expect to achieve the following objectives: 1) a growing database of flux-tower based, continuous CO_2 mixing ratio observations suitable for application to continental inversions; 2) continued development and evaluation of a comprehensive analysis system for estimation of monthly CO_2 exchange across North America at high spatial resolution; 3) significant reduction in the uncertainty in the annual net North American CO_2 flux and its interannual variations as compared to previously published estimates; 4) spatially and temporally resolved terrestrial CO_2 fluxes and uncertainty estimates for 2004 through 2008 encompassing all of North America; 5) quantitative evaluation of the spatial and temporal coherence of the data provided by the North American flux observation network.

4.3 Relevance to Call

The research proposed here directly addresses **solicited area C** on page 2 of the program announcement: *"using empirical data, synthesized datasets, existing models, data assimilation techniques, and theory to advance the ability to quantify spatial patterns and variability of carbon sources and sinks between the atmosphere-land at regional to global scales ... and improve future climate predictions by incorporating a dynamic understanding of the carbon cycle into models."*

4.4 Data

4.4.1 Flux measurements

The micrometeorological approach known as eddy covariance is an effective method of direct observation of NEE and has been successfully applied to long-term observation of NEE of CO_2 at many terrestrial sites. Continuous NEE observations are currently implemented at more than 200 sites worldwide (Baldocchi et al, 2001). Surface-layer, tower-based eddy covariance measures the net flux of CO_2 (and energy, momentum) an area of order 1 km^2. Up-scaling the absolute value of flux measurements is challenging because of both the spatial complexity of terrestrial landscapes (e.g. soils and topography, vegetation cover, land use and land use history) and concerns about systematic errors in long-term eddy-covariance data caused by stability dependent effects such as drainage flows. The flux towers have proven, however, to be excellent tools for examining seasonal and interannual variability in NEE of CO_2 (e.g. Goulden et al, 1998; Ciais et al., 2005; Ricciuto et al, in press). The first long-term measurements were initiated in 1991 (Wofsy et al., 1993). Rapid network expansion occurred in the late 1990s.

Figure 1 shows current sites operating in N. America according to the AmeriFlux web site. The density of sites is sufficient to encompass the principal climatic regions of North America, and provide redundancy within important plant functional types. Most U.S. sites report their processed flux measurements periodically to the DOE's CDIAC program, where the data are publicly archived. Use of the data is allowed as long as a generous fair-use policy is followed. Thus the network provides an effective and operational base of information concerning ecosystem-atmosphere carbon fluxes across the continent. One of the PIs (Davis) is a flux tower PI and member of the AmeriFlux science steering group, thus is very cognizant of the processes required to work with AmeriFlux and Fluxnet Canada data.

4.4.2 Mixing ratio measurements

This project builds upon the NOAA Global Monitoring Division (GMD) network of flask measurements (e.g. Conway *et al.,* 1994), aircraft profiles and tall towers (e.g. Bakwin *et al.,* 1998), and enhances this network with high-quality CO_2 mixing ratio measurements on 13 AmeriFlux towers, 5 of whose CO_2 instrumentation will be maintained via funds requested here, and all of whose mixing ratio data will be sought to make a uniform data product via this project. We will also work with 3 Fluxnet Canada sites with similar data. Note that most flux towers do not maintain CO_2 measurements of sufficient absolute accuracy or long-term precision to be useful in atmospheric inversion studies. These data will be further complemented by a mountaintop network in the Rockies (Stephens, pers. comm.).

4.4.2.1 Flask, tall tower, and aircraft profile data:

The global flask network, basis for numerous previous studies of the global carbon budget, remains a critical backbone of data that will be utilized in this study. The flasks, collected weekly at sites distributed around the globe and located primarily in the marine boundary layer at coastal locations, provide an important background data source for our studies. The flask data provide global coverage, temporal continuity (in some cases lasting for decades), and meticulous calibration to absolute standards. One such flask collection site, the WLEF tall tower in northern Wisconsin (Bakwin *et al.,* 1998), will be an important intercalibration point in the flux tower CO_2 network.

Newer observations that will be utilized in this project include continuous tall tower data and periodic aircraft profiles. NOAA GMD initiated continuous, well-calibrated CO_2 mixing ratio measurements traceable to World Meteorological Organization (WMO) primary standards on very tall communications towers in 1991 (Bakwin et al., 1998). From 1991 through 2002, only one to two sites were operated, including the WITN tower in North Carolina (1991 – 1999), the WLEF tower in Wisconsin (1994 – present), and the KWKT tower in Texas (2001 – present). An expansion of this network is currently establishing a number of additional tall tower sites. The locations of existing and planned tall towers are shown in Figure 1. The number of new sites to be instrumented and the timing of the expansion is unfortunately uncertain due to recent funding cuts, but the sites shown in Figure 1 have a high probability of being instrumented in the coming 1-2 years. The tall tower data have the advantage that the top level where data is collected, typically 400 m to 500 m above ground usually remains above the nocturnal inversion (Yi et al, 2001).

The NOAA GMD group has also initiated periodic airborne sampling of CO_2. These airborne profiles begin in the continental boundary layer and extend up to several kilometers above ground, and currently utilize flask-sampling technology. NOAA GMD is in the process of expanding this network pending resolution of recent budgetary issues. Sites where sampling is currently underway are shown in Figure 1. The current sampling frequency is roughly every other week.

4.4.2.2 AmeriFlux tower CO_2 mixing ratio observations:

Recent support from the Department of Energy funded installation of highly accurate CO_2 mixing ratio systems at five existing AmeriFlux sites including Canaan Valley, WV, Chestnut Ridge, TN, the WLEF tall tower near Park Falls, WI, and sites in Missouri and Montana (Fig 1). In addition, separate funding has enabled several additional AmeriFlux and Fluxnet Canada sites to implement well-calibrated CO_2 measurements (Fig 1). All are intended to be long-term CO_2 mixing ratio measurement sites that will be used in atmospheric inversions. This proposal seeks funds to support in part the maintenance and intercalibration of this network, and data intepretation for all AmeriFlux CO_2 sites.

The surface layer mixing ratios measured at these towers, when subsampled for midday conditions, are very similar to the mixing ratio of the mixed layer (e.g. Yi *et al.,* 2004). Butler *et al.,* (in preparation) shows that further, the small difference between the surface layer mixing ratio and the mid-convective boundary layer (CBL) can be estimated from micrometeorological scaling arguments that have been fitted to the CO_2 flux and mixing ratio measurements from the 447 m tall WLEF tower. The average bias for hourly data is less than 0.2 ppm in summer, less than 0.1 ppm in spring and fall, and less than 0.5 ppm in winter (when mixing is the weakest). The average annual bias for houly data is less than 0.05 ppm. Data from the surface layer, subsampled for midday conditions, contain abundant large-scale synoptic and seasonal structure (e.g. Bakwin *et al.,* 2004; Hurwitz *et al.*, 2004). Nocturnal data are strongly influenced by local stability and more difficult to interpret in larger scale analyses. The absence of data that often reaches above the nocturnal boundary layer, characteristic of tall towers, is a disadvantage of the virtual tall tower approach.

Obtaining high-quality CO_2 mixing ratio data on flux towers requires care. Penn State has developed an instrument, in collaboration with Britt Stephens at NCAR, based on the LI-820 sensor. The instrument is pressure and temperature controlled, the sample air and calibration gases brought to a common and very low dew point temperature, and the system is calibrated with four WMO-traceable standard gases every two hours. A single target tank is measured one every 30 minutes. Side-by-side laboratory and field tests conducted both at Penn State and NCAR suggest that these systems can attain precision and accuracy of 0.3 ppm or better in the field.

As a further means to evaluate the accuracy of the systems, this earlier design of the system was deployed for several months at the WLEF tower in Park Falls, WI, where a NOAA GMD system also measured CO_2 mixing ratio. The NOAA and PSU systems had independent filtering and, more importantly, independent drying. In addition, the NOAA system used a LI-6251. The difference between the daily mean PSU value and the daily mean NOAA value was consistently less than ±0.3 ppm. The accuracy of the improved systems at AmeriFlux towers should reach ±0.2 ppm CO_2. This sensor design has been propagated to additional AmeriFlux sites (Indiana, Nebraska, Oregon), and is nearly the same as the Rocky Mountain network instrument.

Intercalibration of the network, beyond the WMO-primary standard traceable working calibration gases at each site, will be provided by permanent co-location of one of the PSU instruments with a NOAA tall tower (currently the WLEF tower, WI) and a long-term archive CO_2 tank at each site, sampled once a day and remaining at each site for a decade or more. In addition, we have budgeted for travel in year 2 to 5 flux towers with well-calibrated CO_2 measurements. We will bring a mobile instrument and independent calibration gases and run alongside these sites as another means of intercalibration.

4.5 Models

4.5.1 The Simple Biosphere Model (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). SiB has been coupled to the Regional Atmospheric Modeling System (RAMS) and used to study PBL-scale interactions among carbon fluxes, turbulence, and CO_2 mixing ratio (Denning et al., 2003) and regional-scale controls on CO_2 variations (Nicholls et al., 2004; Corbin, 2005; Wang et al, 2006). 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_2. 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, submitted). Under separate support from NASA and in collaboration with James Collatz, 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 are addressing this problem by developing and testing a prognostic phenology module for SiB (and for the Community Land Model, CLM). We are assimilating vegetation imagery into the prognostic phenology model to estimate its parameters (e.g., growing degree day thresholds), rather than forcing it with the satellite data. A new project supported by DOE-NICCR will support development and testing an explicit treatment of phenology and physiology of agricultural crops, and parameterization of the crop model using extensive agricultural databases.

4.5.2 Parameterized Chemical Transport Model (PCTM)

The parameterized chemistry and transport model (PCTM, Kawa et al, 2004) used for global CO_2 simulations was adapted from an established full-chemistry/transport model at NASA Goddard Space Flight Center (Douglass and Kawa,1999; Douglass et al., 2003). At the core of this CTM is the transport code of Lin and Rood (1996), which is formulated in flux form and adopts a semi-Lagrangian algorithm. It is driven by analyzed meteorological fields from NASA's Goddard Earth Observation System, Version 4 (GEOS-4) data assimilation system (DAS). The cycling of the GEOS-DAS is in six-hour windows, using the observations within ±3 hours of the analysis time and a six-hour forecast. Meteorological fields are output from the GEOS-DAS and input to the CTM every 6 hours, including the cloud-mass fluxes and turbulence parameters necessary to drive the CTM.

Kawa et al (2004) evaluated this model by performing forward simulations driven by fossil fuel emissions, air-sea gas exchange, and annually-balanced but seasonally-varying terrestrial NEE (as specified in the TransCom 3 experiment, Gurney et al, 2002). They found that the model reproduced seasonal variations quite well relative to flask observations. We have performed a 5-year simulation of terrestrial photosynthesis and ecosystem respiration driven by GEOS-DAS surface weather on a 1º x 1.25º global grid, and prescribed hourly values of simulated NEE as boundary forcing to the PCTM. The resulting 1º x 1.25º simulations of atmospheric CO_2 were compared to observed variations at six continuous measurement sites in the US and Canada. Seasonal and synoptic variations are reasonably well captured by the model, though nighttime maxima are not strong enough at many locations. This reflects both the relatively coarse vertical resolution of the PCTM near the surface and the six-hourly frequency of the turbulence statistics derived from the analyzed weather.

4.5.3 Lagrangian Particle Dispersion Model (LPDM)

The LPDM (Uliasz and Pielke, 1991; Uliasz, 1993, 1994; Uliasz et al., 1996) accounts for transport by resolved advection and 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_2 and the lateral boundary conditions, though with sufficient integration time the former become negligible.

We have developed a method for regional CO_2 flux inversion using the LPDM driven by the analyzed weather and a first-guess of surface fluxes produced by SiB-CASA. The method involves four steps: (1) forward simulation of photosynthesis, respiration, and decomposition using SiB-CASA driven by surface weather analyses; (2) calculation of a large number of backward-in-time particle trajectories from each observation point ("receptor") in space and time, driven by the 3-D transport fields from the RUC archive; (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 has been tested using synthetic data at both continental scale and in the Ring of Towers experiment, with LPDM driven by output from RAMS instead of RUC analyses.

4.5.4 NOAA Rapid Update Cycle (RUC) analyses

The Rapid Update Cycle (RUC, Benjamin et al, 2004a,b) is the only 1-hour assimilation and mesoscale forecast cycle in the world running as part of an operational numerical prediction center (US National Centers for Environmental Prediction). In June 2005, the horizontal resolution of the operational RUC was changed from 20 km to 13 km. The analysis is made on 50 levels in the vertical, with very high resolution in the planetary boundary layer (PBL). This facilitates representation of near-surface gradients in CO_2 for the interpretation of tower CO_2 data.

With separate support from NOAA ESRL, we have been archiving surface weather and transport fields (winds, turbulence, cloud mass fluxes) since the beginning of 2006, and have modified the LPDM to generate influence functions for hourly tower CO_2 observations based on these fields. We propose to use the surface weather from the RUC analyses to drive SiB-CASA, producing hourly estimates of GPP and ecosystem respiration on a 13-km grid over the continental US and much of Canada. We will then use the LPDM-derived influence functions to estimate spatially-explicit biases in GPP and in carbon storage pools in SiB-CASA that are consistent with the CO_2 observations as described in section 4.6.2 below.

4.6 Analyses

Testing our hypotheses requires assimilation of both AmeriFlux tower eddy-covariance CO_2 flux and atmospheric CO_2 observations as described in section 4.4. We will adopt a Bayesian framework for this analysis and perform a sequential assimilation (Fig 2). First, we will assimilate the flux tower data into the SiB-CASA terrestrial carbon cycle model. This step will deliver prior estimates for surface fluxes of CO_2 in the form of probability distribution functions for model parameters, and analyze space/time covariance in fluxes to constrain the atmospheric inversion. We will focus on a small number of unknown parameters and use very robust optimization algorithms. In the second step, we will assimilate the atmospheric CO_2 concentrations to solve for the terrestrial CO_2 fluxes consistent with both the atmospheric observations and the prior fluxes determined from the flux tower network. The quantities determined by this inversion will include a correction factor for ecosystem productivity, assuming that the flux towers provide a strong prior constraint, and model state variables linked to the respiratory fluxes. We will rely on fast optimization algorithms capable of dealing with the very large number of unknowns that must be determined in this step. We describe these assimilation steps in turn. This sequential approach has the advantage of being able to adjust the method to the specific challenges associated with each assimilation problem. We will then analyze the results of these assimilation steps to address the scientific hypotheses.

4.6.1 Flux tower data assimilation

Our main hypothesis for flux tower data assimilation is that continuous AmeriFlux network CO_2 flux data, when assimilated into the SiB-CASA model, will provide a spatially extensive constraint to the N. American terrestrial carbon cycle, proving particularly valuable at daily to seasonal, as well as interannual time scales. We will test the following sub-hypotheses in pursuit of this overarching hypothesis: (i) A small subset of SiB-CASA model parameters (Prihodko et al, 2006) and state variables (carbon pools, Schaefer et al, submitted) will be sufficient to explain a large fraction of the variance in AmeriFlux CO_2 flux observations when the model is optimized using the flux data. (ii) The likelihood function for the flux tower assimilation is nonconvex (contains several maxima). Assimilation techniques using Gaussian approximations (*e.g.*, the Ensemble Kalman Filter (Evensen, 1994)) will produce biased and overconfident results. (iii) The residuals between modeled and observed terrestrial CO_2 fluxes display statistically significant temporal autocorrelation, extending to seasonal and annual time scales. Failing to account for this autocorrelation causes overconfident and biased parameter and flux estimates. The nature of the temporal autocorrelation will suggest potential modifications to the SiB-CASA model structure.

(iv) The model-observational residual also contains significant spatial coherence. This coherence peaks at seasonal to interannual time scales, and persists within climate-driven biomes (e.g. temperate eastern deciduous forests). (v) Plant functional types provide a sound method for determining coherent terrestrial model parameters. (vi) The assimilation process yields strong constraint on model parameters governing photosynthetic processes, and relatively weak constraints on respiratory parameters. Similarly, the optimized model explains daily to seasonal variability in observed terrestrial CO_2 fluxes much better than it explains variability in fluxes aggregated over annual to interannual time scales. We will assimilate the flux tower data in three main steps, then extrapolate these results over space to provide input to the atmospheric inversion step (Fig 2). A description of the steps in the flux tower assimilation follows:

4.6.1.1 Assimilation methodology

(1) We will identify key model parameters that dominate the model response. A full Bayesian inversion for all parameters of the SIB-CASA model is infeasible given the available computational resources. This is due the relatively large CPU requirement for a single function evaluation and the geometric increase of the necessary function evaluations as the number of parameters increases (Bellman, 1961). This problem can be addressed either via faster (though less numerically robust) numerical techniques such as an ensemble Kalman filter or adjoint, or by reducing the dimensionality of the problem (*e.g.*, Rayner *et al*, 2005; Sachs *et al.*, 2006). For flux tower data assimilation we will work to retain a global optimization method, and to reduce the dimensionality of the problem via an objective model pruning technique. Specifically, we will derive prior probability density function for all model parameters from the published literature and then assess which parameters result in the largest change in the posterior likelihood. We will use a fractional factorial approach as it requires only a limited number of model runs while allowing the partial consideration of parameter interactions (Box *et al.*, 1978). We will select model parameters with the largest variation in the likelihood over the prior range to be estimated in the assimilation step. The remaining parameters will be fixed at their prior estimates.

(2) We will assimilate the flux tower observations. Once the number of model parameters has been pruned to a computationally manageable number, we will adapt a previously developed concept to optimize the model parameters based on flux tower observations using a method that creates a fast approximation of the likelihood function (Knutti et al, 2003). Specifically, we will evaluate the likelihood function at a sparse and space filling set of model parameters using a stratified Latin Hypercube sampling design (Helton and Davis, 2003). We then fit an interpolation surface to the likelihood function using kriging (Cressie, 1993). We will apply a Markov Chain Monte Carlo method that can handle local maxima (Warnes, 2001) to the approximated likelihood function to derive the posterior parameter estimates. The flux tower data will be grouped into a small number of plant functional types consistent with SiB-CASA, and assimilation will be performed to find ecological parameters unique to those plant functional types. Assimilation will also be done for each flux tower individually to search for other potential groupings. The entire 5-year record of flux tower data will be assimilated simultaneously. Carbon storage pools (e.g., live and dead wood, soil carbon) are represented as predicted state variables in SiB-CASA, not parameters, but have a first-order effect on simulated ecosystem respiration. Prior estimates of the size of these pools will be generated by assuming equilibrium conditions (no long term source or sink), but we will experiment with revising the pool sizes for each tower site during the assimilation, reflecting local conditions of disturbance, succession, and management. Unlike model parameters, the pools are not expected to be representative across biomes or plant functional types; if these pools lack coherence across sites, these estimates will become a diffuse prior. Spatially extensive estimates of carbon pools will be obtained from the atmospheric assimilation step.

(3) We will evaluate the difference between modeled and observed terrestrial CO_2 fluxes with two goals in mind; understanding the nature of the model-data difference, and improving the SiB/CASA parameter estimates. We will account for the effects of spatiotemporal autocorrelation of the residuals (*e.g.*, Chevallier et al, 2006; Ricciuto, 2006) on the SiB/CASA parameter estimates by amending the likelihood function. We will use an autoregressive time series model for the temporal autocorrelation (*cf.* Zellner, 1964; Ricciuto, 2006) and a Matern correlation function (Handcock and Stein, 1993) for the

spatial autocorrelation. The Matern class can describe a wide range of correlation functions including the exponential and Gaussian autocorrelation functions. Previous work has shown that neglecting temporal autocorrelation can lead to strongly overconfident and biased parameter estimates (*cf.* Zellner, 1964; Ricciuto, 2006). The parameters describing the spatiotemporal autocorrelation will be jointly estimated with the parameters of the ecosystem model.

(4) Finally, we will compute spatially extrapolated, probabilistic CO_2 flux estimates using the SiB-CASA model and the spatially extensive weather, soils, and vegetation data, and the model parameter joint probability density functions derived in the data assimilation steps. These model realizations will also be used to construct a covariance matrix of prior fluxes to be used in the atmospheric inversion.

4.6.1.2 Hypothesis testing

The following describes, briefly and in sequence, how we will test the sub-hypotheses above.

(i) Once the model pruning and optimization is completed, we will evaluate the degree to which the optimized model explains the flux observations, and break this down as a function of time scale using wavelet analyses (e.g. Braswell *et al*., 2005).

(ii) We will examine the likelihood functions for multiple maxima. We will also compare the assimilation results stemming from our method (which does not require the assumption of a Gaussian shape of the parameter probability function) with results derived from an Ensemble Kalman Filter (Evensen, 1994) which does rely on a Gaussian assumption. Studies focusing on the interannual variations in the terrestrial carbon cycle have shown evidence for local maxima in the likelihood function (*Rayner et al.*, 2005; *Vukicevic et al.*, 2001). We confirmed this (Riccuto, 2006) via assimilation of eddy covariance data into a terrestrial carbon cycle model (TRIFFID, Cox et al, 2000, Matthews et al, 2005). Two maxima were evident in a soil moisture sensitivity parameter, violating a key assumption of the Ensemble Kalman Filter.

(iii) We will compare the results of assimilation with and without the correction for the temporal autocorrelation. We will study the temporal autocorrelation function to see if it suggests any particular processes (e.g. temporal lags in soil hydrology and response to drought stress) that could be improved in the SiB-CASA model structure. If the analyses suggest strong temporal autocorrelation, we will consider breaking up the assimilation process into separate time periods. We will attempt to avoid this, however, and instead reconsider relevant model proceses as being time dependent if suggested by the data.

(iv) We will similarly compare the results of the assimilation with and without the spatial correlation function, and will study the nature of the spatial correlation (e.g. Chevallier *et al*., 2006) in the model-data residual. We will examine the temporal persistence of this spatial correlation, and search for links between spatial correlation and particular processes, model parameters or state variables that could account for this correlation. An initial analysis by Chevallier *et al.,* (2006) found no spatial correlation using multiple flux towers and the ORCHIDEE terrestrial carbon cycle model. Chevallier *et al.,* (2006), however, did not assimilate flux tower data first, and only examined the spatial coherence of daily mean CO_2 fluxes.. We expect that there will be spatially coherent residuals spanning biomes, and associated perhaps with the same processes, poorly represented by SiB-CASA, that are responsible for the seasonal/annual scale temporal autocorrelations that we hypothesize will exist in the residual.

(v) The probability distribution functions of the SiB-CASA parameters determined via joint assimilation of a group of flux towers representing a plant functional type will be compared to the parameters determined by assimilation of each flux tower record independently. If the parameters are not tightly determined in the joint flux tower assimilation, the individual tower data assimilation results will be examined for potential alternative groupings.

(vi) The width of the parameter probability distribution functions (pdfs), and the pdfs of the spatially extrapolated fluxes, will be examined at various temporal and spatial aggregations to determine where the terrestrial CO_2 fluxes are well constrained vs. poorly constrained, and which model parameters/processes are responsible for these spatiotemporal patterns. We will thus quantify the uncertainty in regionally aggregated prior flux estimates at various temporal scales (daily, monthly, annual).

4.6.2 Assimilation of CO_2 Mixing Ratio Data

A fundamental assumption in the two-step assimilation procedure we propose is that high-frequency variations in NEE are driven by radiation and weather and can be successfully modeled by the flux-tower-optimized SiB-CASA. This allows us to accumulate mixing ratio data over a longer period of time to estimate spatial variations in state variables (e.g., carbon stocks) that control the lower frequency source-sink dynamics. We use the model and environmental data to account for spatial and high-frequency time variations of photosynthesis and respiration by assuming that they are driven by well-understood and easily modeled processes (vegetation distribution, 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-CASA. The net ecosystem exchange (NEE) is composed of these two component fluxes:

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

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:

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

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. We 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

$$C_{k,m} = \sum_{i,j,n} \left(\left(\beta_{R,i,j} RESP_{i,j,n} - \beta_{A,i,j} GPP_{i,j,n} \right) C^{*}_{k,m,i,j,n} \right) \Delta t_f \Delta x \Delta y + C_{BKGD,k,m} \qquad \text{(eq 3)}$$

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!). Fossil fuel combustion is specified according to an hourly analysis on a 32-km grid being developed in collaboration with K. Gurney and tested at CSU. 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 Δx and Δy are the sizes of the grid cells in the zonal and meridional direction, and Δt_f is the time step over which the fluxes are applied. The term $C_{BKGD,k,m}$ represents the contribution of "background" CO_2 flowing into the model domain from the larger scales (estimated from the global PCTM analyses). With a little algebra and a healthy dose of computer time, we obtain a simpler representation more practical suitable for optimization:

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

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:

$$C^*_{RESP,obs,cell} = \Delta t_f \Delta x \Delta y \sum_n RESP_{cell,n} C^*_{obs,cell,n}$$
$$C^*_{GPP,obs,cell} = -\Delta t_f \Delta x \Delta y \sum_n GPP_{cell,n} C^*_{obs,cell,n} \qquad \text{(eq 5)}$$

Equation 4 is a linear system which can be written simply as

$$\vec{y} = h\vec{x} \qquad \text{(eq 6)}$$

where $\vec{y}$ is the vector of observations C_{obs} and $\vec{x}$ is the vector of unknown bias terms $\beta_{\text{GPP,cell}}$ and $\beta_{\text{Resp,cell}}$. The Jacobian matrix h contains the influence functions $C^*_{GPP,obs,cell}$ and $C^*_{RESP,obs,cell}$. The rows of h correspond to each observation, and each column corresponds to an unknown bias term β_{RESP} or β_{GPP} at a given grid cell over the 10-day integration period. In practice, we treat the background mixing ratio by prescribing lateral inflow from the global PCTM. We treat errors in this boundary condition additively by augmenting the vector of unknowns $\vec{x}$ with lateral boundary concentrations and "transporting" them to the receptor by augmenting matrix h with additional influence functions for these fluxes.

We minimize a cost function that penalizes model-data mismatch and is regularized by imposing a weak prior constraint:

$$J = (\vec{y} - h\vec{x})^T r^{-1} (\vec{y} - h\vec{x}) + (\vec{x} - \vec{x}_p)^T p^{-1} (\vec{x} - \vec{x}_p)$$

where r is the observation error covariance, and p is the prior error covariance of the unknown β's.

We have implemented this model 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 β's in our case) are generated by the algorithm from an initial forward simulation and calculation of model-data mismatch ($\vec{y} - h\vec{x}$ in our case). An ensemble of forward model integrations (for us, the simple matrix multiplication $h\vec{x}$) 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.

The ensemble yields an approximation of the full error covariance matrix of the β's, the accuracy of which depends on the size of the ensemble. Theoretically, the MLEF estimation approaches the analytical solution when the size of the ensemble is equal to the number of unknowns (this is called the "full-rank" problem). We have verified this behavior for continental and regional inversions of SiB-CASA fluxes by comparing estimates of β (x,y) and its error covariance computed with full-rank ensembles to the analytical solution. The MLEF algorithm includes a strong preconditioning step that reduces the size of ensembles required. In our experiments with synthetic data, we have found that ensembles of 100 members produce results that are almost indistinguishable from the full-rank solution (1800 members).

A key advantage of the estimation of β (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.

We will also experiment with a generalization of the above assimilation procedure in which we vary not only the two bias terms for GPP and ecosystem respiration, but also the organic pools in SiB-CASA with intermediate turnover times: live wood, coarse woody debris, and slow soil carbon. Faster pools adjust quickly to changes in these, and the slowest soil carbon pools have very little influence on respiration over the time scales of the project. The advantage of this approach is that these pools change much more slowly than the 10-day assimilation cycle envisaged above, allowing much more data to be brought to bear on their values. Estimation of woody biomass and soil carbon can also be evaluated independently by in-situ data, and are updated by the model logic itself rather than reset time and again by the assimilation process. The disadvantage is that the fully nonlinear SiB-CASA must be run inside the MLEF optimization. This is too expensive and slow to run on the 13-km RUC grid, but is certainly feasible on a somewhat coarser resolution ($\Delta x \sim 50 - 100$ km). We will also experiment with estimating multiplicative biases in fossil fuel combustion and space-time variations in errors in lateral boundary conditions specified from the global model.

4.7 Expected Results

The two-stage assimilation procedure we propose will allow us to propagate ecophysiological parameters from local to larger scales without the aggregation errors associated with unjustified assumptions about the representativeness of flux towers. The atmospheric mixing ratio data will allow us to quantify carbon cycling at lower frequencies and larger scales, yet still remain consistent with the flux network. The use of a model with predictive long-term carbon storage facilitates analysis of sources and sinks due to disturbance, succession, and management in a consistent framework. We will test the hypotheses presented in section 4.2, evaluating the control of climate and other mechanisms over variability in carbon cycling, the consistency of the flux and atmospheric observations, the representativeness of site studies, and the bottom-up inventory studies.

4.8 Timeline/Management

In year one, we will establish the site calibration procedures, set up the web site for distribution of mixing ratio time series, and couple the RUC and SiB-CASA analyses. In year two, we will continue the site maintenance, distribute the observations, assimilate the flux tower record into SiB-CASA, and begin the atmospheric data assimilation. In year three, we will analyze the covariance structure, perform the atmospheric data assimilation, and publish our results. Richardson, and Miles will be responsible for developing mixing ratio calibration and intercomparison methods, archiving the mixing ratio data, and distributing them to the public. Davis, Keller and Denning will be responsible for the two-stage assimilation of flux and mixing ratio observations into SiB-CASA, and for the analysis of error covariance. All investigators will participate in the analysis and interpretation of the results.

4.9 References Cited

Baker, D. F., R. M. Law, K. R. Gurney, P. Rayner, P. Peylin, A. S. Denning, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, K. Masarie, M. Prather, B. Pak, S. Taguchi, and Z. Zhu, TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988-2003, *Global Biogeochemical Cycles*, **20**, GB1002, doi:10.1029/2004GB002439, 2006.

Baker, I.T., A.S. Denning, N. Hanan, L. Prihodko, P.-L. Vidale, K. Davis and P. Bakwin, Simulated and observed fluxes of sensible and latent heat and CO_2 at the WLEF-TV Tower using SiB2.5, *Global Change Biology*, **9**, 1262-1277, 2003.

Bakwin, P. S., P. P. Tans, D. F. Hurst, C. L. Zhao, Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program., *Tellus Series B-Chemical and Physical Meteorology*, **50**(5), 401-415, 1998.

Bakwin, P. S., K. J. Davis, C. Yi, S. C. Wofsy, J. W. Munger, L. Haszpra, Z. Barcza, Regional carbon dioxide fluxes from mixing ratio data, *Tellus Series B-Chemical and Physical Meteorology*, **56**(4), 301-311, 2004.

Baldocchi, D., E. Falge, L. H. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, J. Fuentes, A. Goldstein, G. Katul, B. Law, X. H. Lee, Y. Malhi, T. Meyers, W. Munger, W. Oechel, K. T. P. U, K. Pilegaard, H. P. Schmid, R. Valentini, S. Verma, T. Vesala, K. Wilson, and S. Wofsy, 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**, 2415-2434, 2001.

Bellman, R., *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, NJ, 1961.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, R. Bleck, Mesoscale weather prediction with the RUC hybrid isentropic—terrain-following coordinate model, *Monthly Weather Review*, **132**(2), 473-494, 2004.

Benjamin, S. G., D. Devenyi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, T. L. Smith, G. S. Manikin, An hourly assimilation-forecast cycle: The RUC, *Monthly Weather Review*, **132**(2), 495-518, 2004.

Box, G.E.P., W.G. Hunter, and J.S. Hunter, *Statistics for experimenters*, John Wiley and Sons, New York, 1978.

Braswell, B.H., W.J. Sacks, E. Linder, and D.S. Schimel, Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, *Global Change Biology*, **11**, 335-355, 2005.

Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais, On the assignment of prior errors in Bayesian inversions of CO_2 surface fluxes, *Geophysical Research Letters*, **33**, L13802, 2006.

Ciais, P., M. Reichstein, N. Niovy, A. Granier, J. Ogee, V. Allard, M. Aubinet, N. Buchmann, C. Bernhofer, A. Carrara, F. Chevallier, A. D. Friend, P. Friedlingstein, T. Grunwald, B. Heinesch, P. Keronen, A. Knohl, G. Krinner, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J.-M. Ourcival, D. Papale, K. Pilegaard, S. Rambal, G. Seurfert, J. F. Soussana, M. Sanz, E. D. Schultze, T. Vesala, R. Valentini, Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, **437**, 530-533, doi:10.1038/nature03972, 2005.

Collatz, G. J., Ball, J. T., Grivet, C. and Berry, J. A., 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, 1991.

Collatz, G. J., Ribas-Carbo, M. and Berry, J. A., Coupled photosynthesis-stomatal conductance model for leaves of C_4 plants, *Aust. J. Plant Physiol.*, **19**, 519-538, 1992.

Conway, T. J., P. P. Tans. L. S. Waterman, K. W. Thoning, Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration Climate Monitoring and Diagnostics Laboratory global air sampling network, *Journal of Geophysical Research-Atmospheres* , **99**(D11), 22831-22855, 1994.

Corbin, *Evaluating Spatial, Temporal, and Clear-Sky Errors in Satellite CO_2 Measurements*, M. S. Thesis, Department of Atmospheric Science, Colorado State University, 183 pp, 2005.

Cox, P.M., R.A. Betts, C.D. Jones, S.A. Spall, and I.J. Totterdell, Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, **408**, 184-187, 2000.

Cressie, N.A.C., *Statistics for spatial data*, Wiley Intersciences, New York, 1993.

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, The common land model (CLM), *Bulletin of the American Meteorological Society*, **84**, 1013–1023, 2003.

Davis, K. J., P. S. Bakwin, C. X. Yi, B. W. Berger, C. L. Zhao, R. M. Teclaw, and J. G. Isebrands, The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, *Global Change Biology*, **9**, 1278-1293, 2003.

Denning, A. S., I. Y. Fung, and D. Randall, Latitudinal gradient of atmospheric CO_2 due to seasonal exchange with land biota, *Nature*, **376**, 240-243, 1995.

Denning, A.S., J.G. Collatz, C. Zhang, D.A. Randall, J.A. Berry, P.J. Sellers, G.D. Colello and D.A. Dazlich, Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. Part 1: Surface carbon fluxes, *Tellus*, **48B**, 521-542, 1996a.

Denning, A.S., D.A. Randall, G.J. Collatz and P.J. Sellers, Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. Part 2: Spatial and temporal variations of atmospheric CO_2, *Tellus*, **48B**, 543-567, 1996b.

Denning, A.S., T. Takahashi and P. Friedlingstein, Can a strong atmospheric CO_2 rectifier effect be reconciled with a "reasonable" carbon budget? *Tellus*, **51B**, 249-253, 1999.

Denning, A.S., et al., Simulated and observed variations in atmospheric CO_2 over a Wisconsin forest using a coupled Ecosystem-Atmosphere Model, *Global Change Biology*, **9**, 1241-1250, 2003.

Denning, A.S. et al., *Science Implementation Strategy for the North American Carbon Program*. Prepared for the Carbon Cycle Science Steering Group and the Interagency Working Group On Carbon. Available at http://www.carboncyclescience.gov, 2005.

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.

Douglass, A. R. and S. R. Kawa, Contrast between 1992 and 1997 high-latitude spring Halogen Occultation Experiment observations of lower stratospheric HCl, *Journal of Geophysical Research-Atmospheres*, **104**(D15), 18739-18754, doi: 10.1029/1999JD900281, 1999.

Douglass, A. R., M. R. Schoeberl, R. B. Rood, S. Pawson, Evaluation of transport in the lower tropical stratosphere in a global chemistry and transport model, *Journal of Geophysical Research-Atmospheres*, **108**(D9), 4259, doi:10.1029/2002JD002696, 2003.

Engelen, R. J., A. S. Denning, and K. R. Gurney, On error estimation in atmospheric CO_2 inversions, *Journal of Geophysical Research*, **107**, 4635, doi:10.1029/2002JD002195, 2002.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, *Journal of Geophysical Research-Oceans*, **99**(C5), 10143-10162, 1994.

Farquhar, G. D., S. V. Caemmerer, and J. A. Berry, A Biochemical-Model of Photosynthetic CO_2 Assimilation in Leaves of C-3 Species, *Planta*, **149**, 78-90, 1980.

Fletcher, S. J. and M. Zupanski, A framework for data assimilation which allows for non-Gaussian errors, submitted to *Proc. Royal Soc. of London*.

Friedlingstein, P., P. M. Cox, R. A. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. C. Doney, M. Eby, I. Y. Fung, G. Bala, J. John, C. D. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, N. Zeng, Climate-carbon cycle feedback analysis: Results from the C^4MIP model intercomparison, *Journal of Climate,* **19**, 3337-3353, 2006.

Gerbig C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, C. A. Grainger, Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res*., **108**, 4756, doi:10.1029/2002JD003018, 2003a.

Gerbig C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, C. A. Grainger, Toward constraining regional-scale fluxes of CO_2 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, 2003b.

Goulden, M. L., S. C. Wofsy, J. W. Harden, S. E. Trumbore, P. M. Crill, S. T. Gower, T. Fries, B. C. Daube, S. M. Fan, D. J. Sutton, A. Bazzaz, J. W. Munger, Sensitivity of boreal forest carbon balance to soil thaw, *Science*, **279**(5348), 214-217, 1998.

Gurney, K. R., R. M. Law, A. S. Denning, P. J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, S. Fan, I. Y. Fung, M. Gloor, M. Heimann, K. Higuchi, J. John, T. Maki, S. Maksyutov, K. Masarie, P. Peylin, M. Prather, B. C. Pak, J. Randerson, J. Sarmiento, S. Taguchi, T. Takahashi, and C. W. Yuen, Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models. *Nature*, **415**, 626-630, 2002.

Hanan, N. P., J. A. Berry, S. B. Verma, E. A. Walter-Shea, A. E. Suyker, G. G. Burba, and A. S. Denning, Model analyses of biosphere-atmosphere exchanges of CO_2, water and energy in Great Plains tallgrass prairie and wheat ecosystems, *Agricultural and Forest Meteorology*, **131**, 162-179, 2004.

Handcock, M.S., and M.L. Stein, A Bayesian analysis of kriging, *Technometrics*, **35**, 403-410, 1993.

Helliker, B. R., J. A. Berry, A. K. Betts, P. S. Bakwin, K. J. Davis, A. S. Denning, J. R. Ehleringer, J. B. Miller, M. P. Butler, D. M. Ricciuto, Estimates of net CO_2 flux by application of equilibrium layer concepts to CO_2 and water vapor measurements from a tall tower, *Journal of Geophysical Research-Atmospheres*, **109**, D20106, doi:10.1029/2004JD004532, 2004.

Helton, J.C., and F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering & System Safety*, **81**(1), 23-69, 2003.

Hurwitz, M. D., D. M. Ricciuto, P. S. Bakwin, K. J. Davis, W. G. Wang, C. X. Yi, and M. P. Butler, Transport of carbon dioxide in the presence of storm systems over a Northern Wisconsin forest, *Journal of the Atmospheric Sciences*, **61**, 607-618, 2004.

Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting, On aggregation errors in atmospheric transport inversions, *J. Geophys. Res*., **106**, 4703-4715, 2001.

Kawa, S. R., D. J. Erickson III, S. Pawson, Z. Zhu, et al, Global CO_2 transport simulations using meteorological data from the NASA data assimilation system, *Journal of Geophysical Research-Atmospheres*, **109**(D18), 18312, doi:10.1029/2004JD004554. 2004.

Knutti, R., T.F. Stocker, F. Joos, and G.K. Plattner, Probabilistic climate change projections using neural networks, *Climate Dynamics*, **21**(3-4), 257-272, 2003.

Law, R. M., P. J. Rayner, L. P. Steele, I. G. Enting, Data and modelling requirements for CO_2 inversions using high-frequency data, *Tellus Series B-Chemical and Physical Meteorology*, **55**(2), 512-521, 2003.

Lin, S.-J. and R. B. Rood, Multidimensional flux-form semi-lagrangian transport schemes, *Monthly Weather Review*, **124**, 2046-2070, 1996.

Matthews, H.D., M. Eby, A.J. Weaver, and B.J. Hawkins, Primary productivity control of simulated carbon cycle-climate feedbacks, *Geophysical Research Letters*, **32**(14), 2005.

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.

Nicholls, M.E., A.S. Denning, L. Prihodko, P.-L. Vidale, K. Davis, P. Bakwin, A multiple-scale simulation of variations in atmospheric carbon dioxide using a coupled biosphere-atmospheric model, *Journal of Geophysical Research,* **109**, D18117, doi:10.1029/2003JD004482, 2004.

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_2 surface fluxes from atmospheric trace gas observations, *J. Geophys. Res*., **110**, D24304, doi:10.1029/2005JD006157, 2005.

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_2 flux variations over land and oceans, *Global Biogeochem. Cycles*, **19**, GB1011, doi:10.1029/2003GB002214, 2005.

Prihodko, L., A. S. Denning, N. P. Hanan, I. Baker, K. Davis, Sensitivity, uncertainty and time dependence of parameters in a complex land surface model, *Agric. and Forest Meterol.*, in press.

Randall, D.A., P.J. Sellers, J.A. Berry, D.A. Dazlich, C. Zhang, C.J. Collatz, A.S. Denning, S.O. Los, C.B. Field, I. Fung, C.O. Justice and C.J. Tucker, A revised land surface parameterization (SiB2) for atmospheric GCMs. Part 3: The greening of the CSU GCM. *J. Clim.*, **9**, 738-763, 1996.

Randerson, J. T., M. V. Thompson, T. J. Conway, I. Y. Fung, and C. B. Field, The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cycles*, **11**, 535-560, 1997.

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.

Ricciuto, D.M., *Diagnosing uncertainty and improving predictions of terrestrial CO_2 fluxes at multiple scales through data assimilation*, Ph.D. dissertation, The Pennsylvania State University, 2006.

Ricciuto, D. M., M. P. Butler, K. J. Davis, B. D. Cook, P. S. Bakwin, A. Andrews, R. M. Teclaw, Determining the causes of interannual variability in ecosystem-atmosphere carbon dioxide exchange in a northern Wisconsin forest using a Bayesian synthesis inversion, *Agriculture and Forest Meteorology*, in press.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann, Time dependent atmospheric CO_2 inversions based on interannually varying tracer transport, *Tellus*, **B55**, 488–497, 2003.

Schaefer, K., A.S. Denning, N. Suits, Joerg Kaduk, I. Baker, S. Los, and L. Prihodko, The effect of climate on inter-annual variability of terrestrial CO_2 fluxes, *Global Biogeochemical Cycles,* **16**, 1102, doi:10.1029/2002GB001928, 2002.

Schaefer, K., A. S. Denning, and O. Leonard, The winter Arctic Oscillation, the timing of spring, and carbon fluxes in the northern hemisphere, *Global Biogeochemical Cycles,* **19**, GB3017, doi:10.1029/2004GB002336, 2005.

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*.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher, A simple biosphere model (SiB) for use within general circulation models, *J. Atmos. Sci.*, **43**, 505-531, 1986.

Sellers, P.J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, L. Bounoua, A revised land surface parameterization (SiB2) for atmospheric GCMs, Part 1: Model formulation. *J. Clim.*, **9**, 676-705, 1996a.

Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, D. A. Randall, 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, 1996b.

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.

Uliasz, M. and R.A. Pielke, 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, 1991.

Uliasz, M., The atmospheric mesoscale dispersion modeling system (MDMS). *J. Appl. Meteor.*, **32**, 139-149, 1993.

Uliasz, M.,Lagrangian particle modeling in mesoscale applications. *Environmental Modelling* II, ed. P. Zannetti, Computational Mechanics Publications, 71-102, 1994.

Uliasz, M., R.A. Stocker, and R.A. Pielke, Regional modeling of air pollution transport in the southwestern United States, *Environmental Modelling III*, ed. P. Zannetti, Computational Mechanics Publications, 145-182, 1996.

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.

Vukicevic, T., B.H. Braswell, and D. Schimel, A diagnostic study of temperature controls on global terrestrial carbon exchange, *Tellus Series B-Chemical and Physical Meteorology*, **53**(2), 150-170, 2001.

Wang, J.-W., *Observations and simulations of synoptic, regional, and local variations of atmospheric CO_2*, M. S. Thesis, Colorado State University, 146 pp, 2005.

Wang, J.-W., A. S. Denning, L. Lu, I. T. Baker, K. D. Corbin, K. J. Davis, Observations and simulations of synoptic, regional, and local variations in atmospheric CO_2, submitted to *J. Geophys. Res. Atmos.*

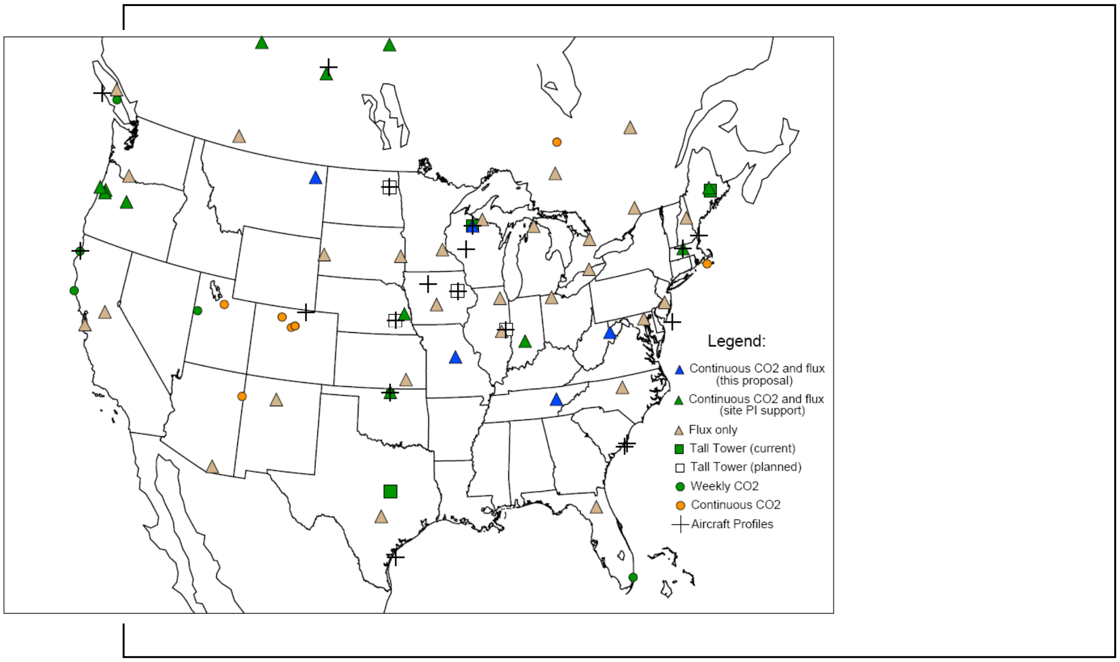
Warnes, G.R., The Normal Kernel Coupler: An adaptive Markov Chain Monte Carlo method for efficient sampling from multi-modal distributions, in *University of Washington, Department of Statistics Working Papers, Report No. 39*, 2001.

Wofsy, S. C., M. L. Goulden, J. W. Munger, S. M. Fan, P. S. Bakwin, B. C. Daube, s. L. Bassow, F. A. Bazzaz, Net exchange of CO_2 in a midlatitude forest, *Science*, **260**(5112), 1314-1317, 1993.

Wofsy, S.C. and R.C. Harriss, The North American Carbon Program (NACP), Report of the NACP Committee of the U.S. Carbon Cycle Science Program. Washington, D.C. US Global Change Research Program. Available at: http://www.esig.ucar.edu/nacp/, 2002.

Yi, C. X., K. J. Davis, B. W. Berger, P. S. Bakwin, Long-term observations of the dynamics of the continental planetary

boundary layer, *Journal of the Atmospheric Sciences* , **58**(10), 1288-1299, 2001.
Yi, C., K. J. Davis, P. S. Bakwin, A.S. Denning, N. Zhang, A. Desai, J. C. Lin, and C. Gerbig, The observed covariance between ecosystem carbon exchange and atmospheric boundary layer dynamics at a site in northern Wisconsin, *Journal of Geophysical Research*, **109**, D08302, doi:10.1029/2003JD004164, 2004.
Zellner, A., and G.C. Tian, Bayesian analysis of the regression model with autocorrelated errors, *Journal of the American Statistical Association*, **59**, 763-778, 1964.
Zupanski, M., Maximum likelihood ensemble filter: Theoretical aspects, *Monthly Weather Review*, **133**(6), 1710-1726, 2006.



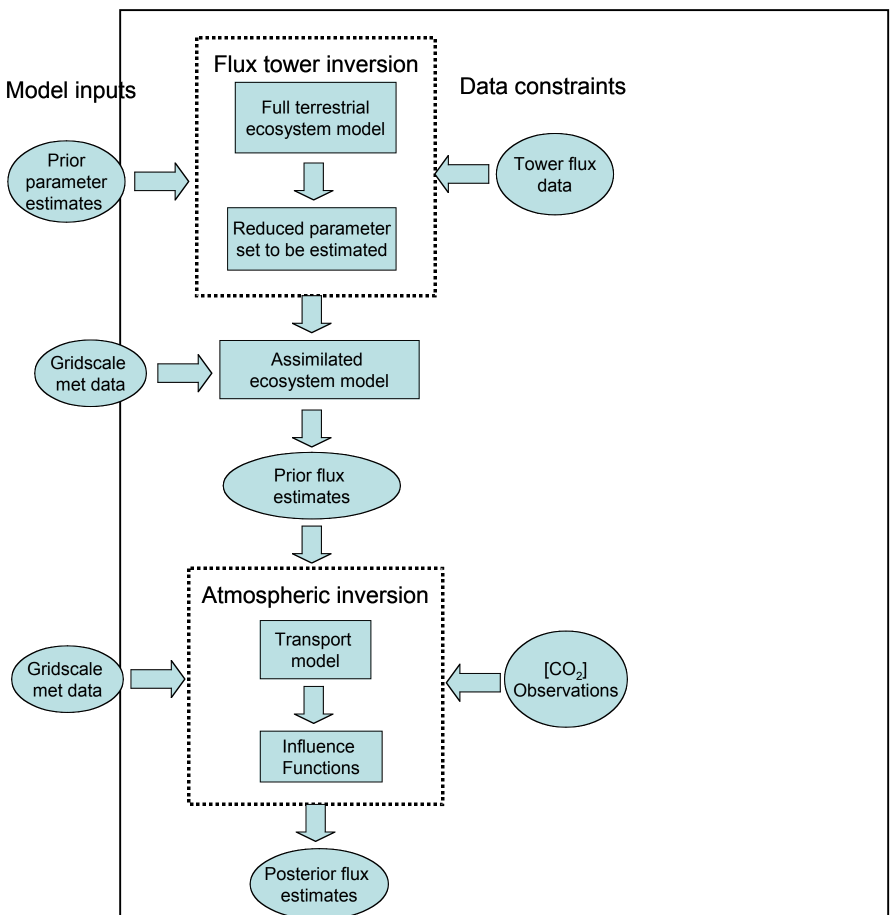


Figure 2: Conceptual overview of the assimilation analyses.