Mesoscale Carbon Data Assimilation for NACP

P.I.: Scott Denning[1], Co-P.I.: Dusanka Zupanski[1]
Co-I's: Milija Zupanski[1], Marek Uliasz[1], G. James Collatz[2]
[1]Department of Atmospheric Science, Colorado State University
[2]NASA Goddard Space Flight Center, Greenbelt, Maryland

Table of Contents

Mesoscale Data Assimilation for NACP

P.I.: Scott Denning[1], Co-P.I.: Dusanka Zupanski[1]

Co-I's: Milija Zupanski[1], Marek Uliasz[1], G. James Collatz[2]

[1]Department of Atmospheric Science, Colorado State University

[2]NASA Goddard Space Flight Center, Greenbelt, Maryland

Abstract

The North American Carbon Program (NACP) is a multi-year program of integrated research supported by many US agencies which seeks to quantify the current budget of CO_2, CO, and CH_4 over North America, to understand and predict the processes governing these fluxes, and to provide timely and practical information products to support management decisions. A major component of NACP is a greatly enhanced system for observing temporal and spatial variations for carbon gases in the atmosphere over North America and adjacent coastal oceans. After 2007, the dense in-situ network of atmospheric measurements for NACP will be augmented by hundreds of thousands of column CO_2 observations each day made from NASA's Orbiting Carbon Observatory (OCO). These new observations are intended both to provide an integral atmospheric constraint to upscaled ("bottom-up") models of carbon exchange processes, and to enable quantitative but process-agnostic estimates of regional monthly sources and sinks by ("top-down") transport inversion. Currently available analytical methods for flux estimation by inverse modeling involve assumptions about the spatial and (especially) temporal patterns of carbon fluxes that will be inappropriate to the much greater density of sampling by NACP and OCO.

We propose to develop a generalized framework for flux estimation from multiple streams of carbon observations, including spectral vegetation and land cover imagery, eddy covariance flux observations, meteorological data, and both in-situ and remotely sensed observations of atmospheric carbon gases. This will be accomplished using Ensemble Data Assimilation (EnsDA) techniques applied to a fully coupled model of regional meteorology, ecosystem carbon fluxes, and biomass burning (SiB-CASA-RAMS). Terrestrial carbon fluxes over North America due to photosynthesis, autotrophic respiration, decomposition, and fires, and a "residual" time-mean source or sink will be simulated by the model. Unknown parameters related to light response, allocation, drought stress, phonological triggers, combustion efficiency, PBL entrainment, convective efficiency, and the time-mean sink will be estimated to obtain optimum consistency with a wide variety of ecological, meteorological, and trace gas observations. The EnsDA method does not require the development of an adjoint of the coupled model, but rather applies an optimization method that involves a large ensemble of forward simulations. Unlike previous high-resolution inversions using transport model adjoint methods, we will not assume surface fluxes remain constant on monthly time scales, and we will not treat the transport model as "perfect." Parameters in the forward coupled model will be quantitatively estimated, as will transport model error. The model will be integrated on a 20-km grid over a domain including most of North America and adjacent oceans, with lateral boundary conditions specified from the output of a global model.

In the first year of the proposed research, we will continue development and local testing of the coupled SiB-CASA model, including new modules for allocation, autotrophic respiration, and decomposition. We will also build a prototype of the EnsDA system using a greatly simplified version of the transport based on a Lagrangian Particle Dispersion Model (LPDM). We will test the EnsDA system in year 2 using synthetic data using the forward coupled model, holding the transport constant and known, and evaluate assumptions about the spatial and temporal covariance of forward model error. In year 3, we will test our prototype EnsDA system with real observations by the mature NACP system, including meteorological data assimilation, transport error estimation, and model improvement. If available, we will also analyze early OCO observations with the EnsDA framework. Finally, we will work with appropriate partners to transfer the EnsDA framework to an operational center for continued analysis and source/sink estimation from available data.

1. Background and Motivation

Only approximately half of current fossil fuel emissions of CO_2 accumulate in the atmosphere, with the remainder sequestered due to uptake by terrestrial ecosystems and the world's oceans (IPCC, 2001). The sink processes modulating today's atmospheric CO_2 increase remain poorly quantified, and future interactions between the carbon cycle, climate, and intentional management now constitute a leading source of uncertainty in projections of 21st century climate change. To address these uncertainties, the U.S. Climate Change Science Program includes an integrated research effort to quantify and understand carbon sources and sinks (Subcommittee on Global Change Research, 2003). An important component of this effort is the North American Carbon Program (NACP), which seeks to address carbon cycle processes at regional to continental scales 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 (Wofsy and Harris, 2002; Denning et al, 2004). ***To address the goals of the NACP, we propose to develop a framework for data assimilation into a process-based regional coupled model of meteorology, terrestrial carbon 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.***

Spatial and temporal variations in the mixing ratio of atmospheric CO_2 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., Enting et al, 1995; Rayner et al, 1999; Gurney et al, 2002, Rödenbeck et al, 2003). Such calculations can provide a critical integral constraint on regional flux estimates that are upscaled from local process understanding using remotely sensed imagery and other spatial data. Transport inversions from currently available atmospheric data inevitably face a trade-off between temporal and spatial aggregation and uncertainties in the estimated fluxes. Monthly mean fluxes over subcontinental-scale regions can be estimated over well-sampled parts of the world to a useful degree of confidence using now-traditional "synthesis inversion" methods (e.g., Gurney et al, 2004, Fig 1), and some insight into the likely impact of model transport error can be approximated by using a growing suite of such codes (Peylin et al, 2002; Gurney et al,

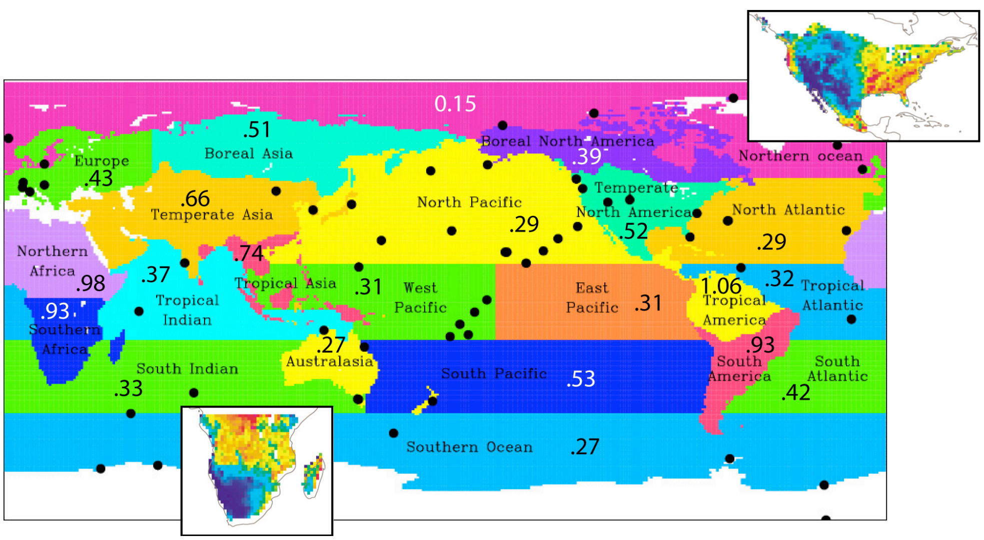


Figure 1: Regions and stations used for the TransCom 3 experiments (Gurney et al, 2002, 2003, 2004). Black circles indicate positions of the 76 flask stations used to estimate surface flux for each region. Insets show examples of the assumed spatial distribution of surface exchange within Temperate North America and Southern Africa regions. Numbers indicate *a posteriori* estimate of uncertainty in annual flux for each region, averaged across 16 participating transport models.

2003).

Synthesis inversion of atmospheric observations involve forward simulation of tracer pulses from regions with prescribed patterns of flux variations in space and time. Prescribing spatial patterns allows other forms of information to be brought to bear on the results of the inverse calculation (e.g., we expect little or no carbon exchange by the Sahara desert or the Greenland ice sheet). If patterns of flux variations are prescribed as *hard constraints* (not adjustable by the optimization procedure) and are incorrect, errors in subregional spatiotemporal variations are inevitably aliased into biases in the estimated fluxes in the regional and time mean (this type of error has been termed "aggregation error," and has been described quantitatively by Kaminski et al, 2001 and Engelen et al, 2002).

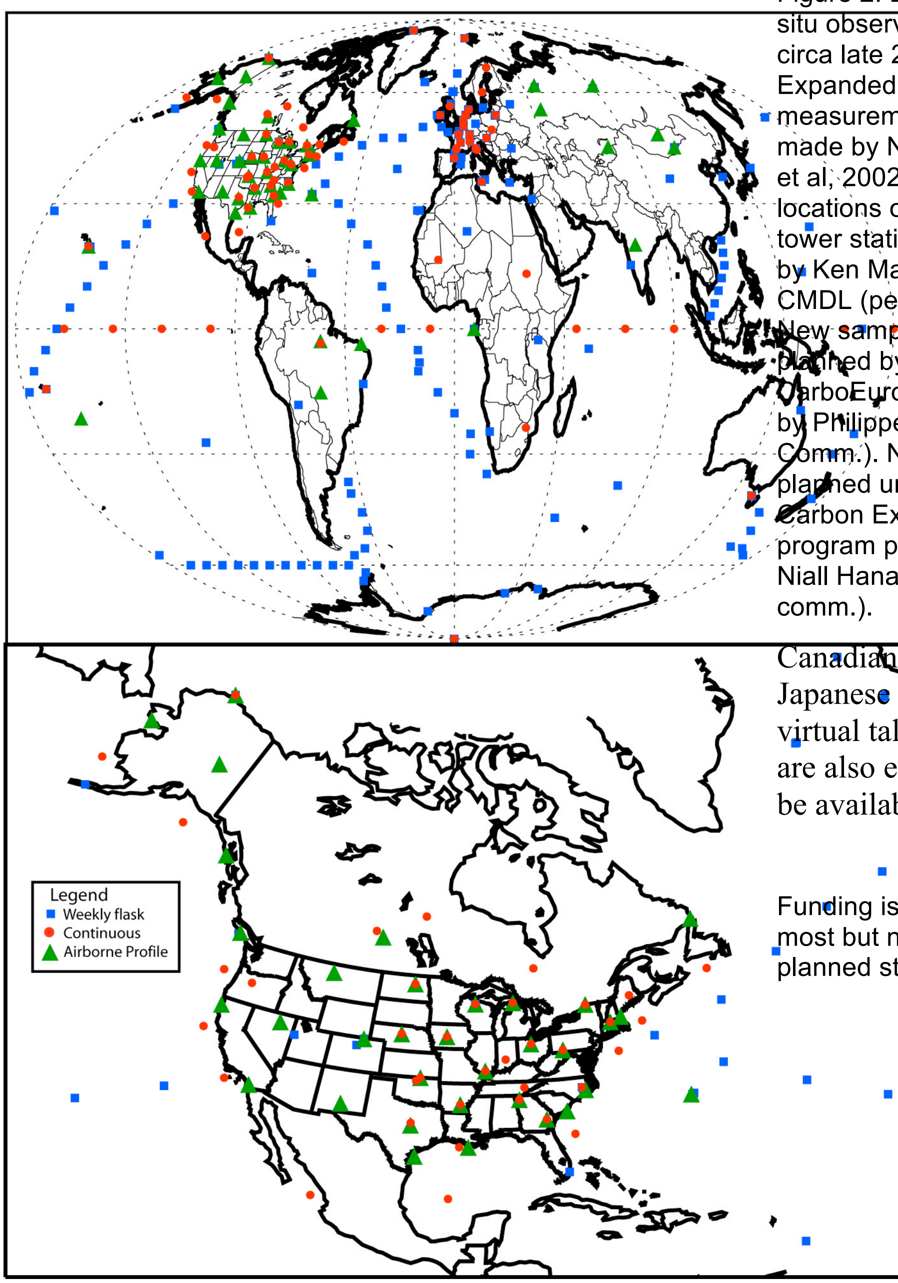


Figure 2: Expected in-situ observing system circa late 2006. Expanded flask measurements to be made by NOAA (Bender et al, 2002), and locations of NACP tall tower stations provided by Ken Masarie, NOAA CMDL (pers. comm.). New sampling locations planned by CarboEurope provided by Philippe Ciais (pers. Comm.). New stations planned under African Carbon Exchange program provided by Niall Hanan (pers. comm.).

Canadian and Japanese sites and virtual tall tower data are also expected to be available.

Funding is in place for most but not all of the planned stations.

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 and Pielke, 1991; Uliasz et al, 1996; Kaminski et al, 1999; Rödenbeck et al, 2003; Gerbig et al, 2003b; Uliasz and Denning, 2004). 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. Rödenbeck et al (2003) used monthly mean CO_2 mixing ratios at dozens of flask stations to estimate fluxes for every grid cell in their global transport model, for example, but uncertainty over most of the world was so high that production of interpretable results required very aggressive post-aggregation to much larger regions. This post-aggregation involved unrealistic assumptions about spatial covariance of surface fluxes: one scenario specified 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. Worse, temporal aggregation errors have scarcely been addressed by inversion studies to date. Rödenbeck et al (2003) aggregate 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, concentrations is negligible (Denning et al, 1995, 1996b). Local observations contradict this assumption, with 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).

The atmospheric observing system is expected to undergo dramatic enhancement in the second half of this decade (Fig 2) as global observing programs (Bender et al, 2002), NACP (Wofsy and Harris, 2002), and a similar effort in Europe (CarboEurope Integrated Project, 2003) deploy additional stations. The density of the enhanced in-situ observing system should enable source/sink estimation to a high degree of confidence over much finer spatial scales than has been possible to date. Continuous measurements of CO_2 (and possibly other relevant gases) from tall towers and coastal buoys, in particular, is expected to dramatically improve the degree of constraint on regional sources and sinks (Law et al, 2002, 2003). Hourly observations exhibit large variations associated with synoptic weather events (Hurwitz et al, 2004) that can be used to estimate upstream fluxes as the fetch changes due to passing weather disturbances (Uliasz and Denning, 2004).

The intent of the enhanced observing system is to provide regional integral constraints for bottom-up source/sink estimates, but extracting the information content of the observations will require new analytical methods. Traditional synthesis inversion involves first generating the Jacobian of the transport operator linking emissions to concentrations (using either forward simulations of flux pulses or adjoint transport from receptors) and then using the observations to solve for all the fluxes in a single step (e.g., Enting, 2002; Rodenbeck et al, 2003; Gurney et al, 2004). Computational considerations limit the temporal and spatial resolution of the fluxes estimated and the observations used by this class of methods, as the size of the matrix to be inverted becomes very large. Estimation of annual mean fluxes for 22 regions with 76 stations in the TransCom 3 calculation, for example, takes much less than a second. Estimation of 12 monthly fluxes for the same 22 regions using mean concentrations at the same 76 stations in 12 months (assuming cyclostationary conditions) takes about a minute. Performing interannual inversion using the TransCom setup requires operations on a matrix of dimension (76 stations x 12 months x 20 years) by (22 regions x 12 months x 20 years), which takes several hours on a very fast CPU. Inverting hourly data at scores of locations over a period of years using this method would increase the size of the matrix by a factor of (24 hours/day x 30 days/month)^2 and would certainly be prohibitively expensive.

Modern data assimilation techniques (e.g., Variational or Kalman filter methods) must be used to reduce the computational dimensions and cost of the inverse problem in a data dense world (Kalnay et al, 2003). 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 reasonably forward model of these fluxes may have temporal coherence on time scales of weeks or even months. This is especially true of coupled data assimilation that can merge multiple streams of data together to obtain the strongest possible constraint on both net carbon exchange and the mechanisms that control it. Kaminski et al (2002) performed a global study of seasonal carbon fluxes using the adjoint of an extremely simple coupled model of terrestrial NPP, ecosystem respiration and atmospheric transport constrained by NDVI imagery, atmospheric CO_2, and eddy covariance data. Their assimilation is being extended to include parameter estimation into a more complex model of terrestrial ecophysiology (Scholze et al, 2003). A regional application of coupled data assimilation to study the carbon cycle of Australia (Wang and Barrett, 2003; Wang and McGregor, 2003) found that carbon fluxes could be estimated with greater confidence, and at higher resolution, using multiple data constraints (ecosystem carbon inventories, satellite vegetation imagery, and hourly atmospheric CO_2) than was possible from either the bottom-up or top-down approaches alone. A significant drawback of coupled carbon assimilation studies to date is that variational or synthesis approaches require the derivation of the adjoint of the coupled model; this is very difficult for complex models and has limited the application of the technique to models with very simplistic process representation.

Using data assimilation methods, it may be possible to estimate the net time-mean surface carbon flux from variations in atmospheric CO_2 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_2 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. Gerbig et al (2003b), for example, estimated regional exchanges with forests and croplands in North America using a very simple model of light-use efficiency and ecosystem respiration constrained by eddy covariance data, vegetation imagery, and airborne CO_2 observations. Though simplistic, the terrestrial flux model was able to represent variations in surface flux on diurnal to synoptic time scales that produced first-order changes in atmospheric mixing ratio, and thus isolate the time-mean component due to unspecified slower processes. Unfortunately, only a few days of anecdotal observations were available, so this prototype study suffered from aggregation error in treating vast areas of cropland and forest as being described by identical control parameters. At a minimum, coupled carbon data assimilation models relying on this separation of time scales to estimate time-mean fluxes from atmospheric CO_2 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 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.

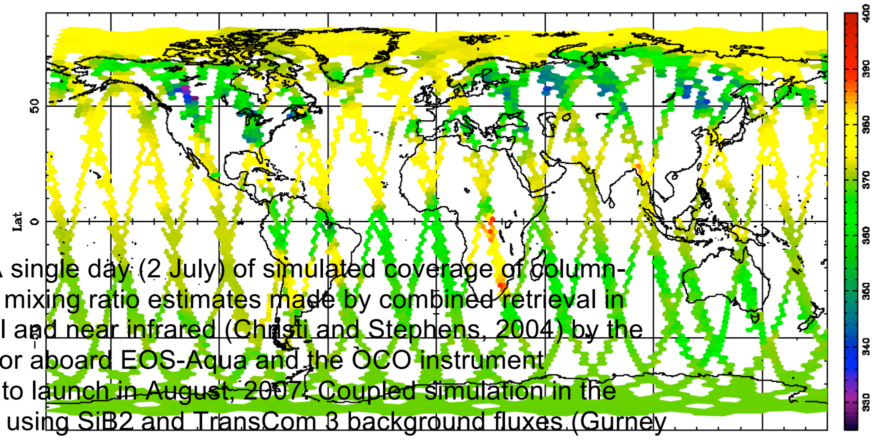


Figure 3: A single day (2 July) of simulated coverage of column-mean CO_2 mixing ratio estimates made by combined retrieval in the thermal and near infrared (Christi and Stephens, 2004) by the AIRS sensor aboard EOS-Aqua and the OCO instrument scheduled to launch in August, 2007. Coupled simulation in the CSU GCM using SiB2 and TransCom 3 background fluxes (Gurney et al, 2003). No cloud screening has been applied.

Beginning in 2007, global observations of column-mean mixing ratio of CO_2 will also be provided by the Orbiting Carbon Observatory (OCO). These data (Fig 3), although more uncertain than the in-situ observations, will dramatically increase the degree of constraint on surface carbon fluxes (Rayner and O'Brien, 2001; Rayner et al, 2002). Estimation of time-mean surface carbon exchanges using satellite CO_2 observations will require appropriate treatment of variations on diurnal and synoptic time scales tha tresult from changes in radiation and weather. Diurnal or cloud/clearsky bias will inevitably result from failure to adequately model these variations, leading to errors in the retrieved fluxes. Coupled data assimilation provides a framework for treatment of these variations.

Even in global inversions constrained by monthly mean observations at predominantly remote sites, transport error is a significant contributor to continental-scale flux uncertainty (Denning et al, 1999; Peylin et al, 2002; Gurney et al, 2003). As enhanced observations enable fluxes to be estimated at finer temporal and spatial resolution, atmospheric transport errors are likely to become more problematic. Global tracer transport models may have insufficient spatial resolution to make the best use of the dense hourly sampling network envisioned under NACP (Fig 2). Currently available global weather analyses are inadequate for driving regional transport models due to insufficient resolution in both time (Δt is typically 3 to 6 hours) and space (Δx is typically 100-250 km), and the standard practice of interpolation to standard pressure levels fails to conserve mass. These products are simply not intended for the purpose of driving quantitative trace gas transport calculations, and using them introduces yet another hard constraint into the flux estimation procedure. Errors in specified transport will unavoidably be aliased into biases in the estimated surface exchanges.

To obtain maximum value from the emerging temporally and spatially dense observing systems, a new approach is needed, in which state-of-the-art data analysis is applied to produce custom meteorological analyses in support of carbon cycle research, rather than relying on standard products intended for weather forecasting. Unlike the weather forecast problem, carbon cycle data assimilation has the luxury of time – weeks or months may elapse after samples are collected before final quality-controlled data are available for analysis. The carbon data assimilation problem also requires very careful treatment of diurnal variability of terrestrial ecophysiology, boundary-layer turbulence, and cumulus convection. In short, we need to control the details of our own modeling environment rather than rely on "off-the-shelf" meteorological products to drive off-line carbon cycle models and inverse calculations. This will require a major research effort to develop both appropriate forward models and assimilation techniques to make use of the wide variety of observations available.

Mesoscale Data Assimilation for NACP

An alternative approach to the variational and adjoint/synthesis techniques 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. These new methods are a unification of ensemble forecasting (e. g., Toth and Kalnay 1993, 1997; Palmer 1993; Molteni et al. 1996) and Kalman filter/smoother (KF/KS) data assimilation methods (Kalman 1960; Jazwinski 1970; Ghil et al. 1981; Cohn 1997). A major advantage of EnsDA methods is that computation of the adjoint of the forward process model is not required. 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.

A common characteristic of all EnsDA approaches is that the optimal state estimate or analysis is sought in an ensemble-spanned subspace, defined by a limited number of forecast model realizations (ensemble members). Ensemble based data assimilation techniques provide a consistent mathematical formalism to updating (cycling) analysis and forecast error covariance matrices, optimally employing information from the observations. This is, therefore, a fully adaptive probabilistic approach to data assimilation and prediction, including the estimates of analysis and forecast *uncertainties* in terms of the analysis and forecast error covariance matrices. Due to the ease of the ensemble framework, a prediction model (e.g., atmospheric, terrestrial, carbon) of any complexity can be used in EnsDA, employing essentially the same algorithm. This substantially reduces the algorithm development effort.

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 a time-mean source or sink of unknown magnitude due to unspecified slower/exogenous processes. The forward coupled model will be built from mature existing components, reducing development time and effort. The forward model will be integrated on a 20-km grid over most of the NACP domain, and will predict the weather, surface carbon exchanges due to photosynthesis, respiration, fossil fuel emissions, and fire. The model will be constrained by satellite land cover, vegetation and fire products, eddy covariance tower data, and atmospheric CO_2 mixing ratios (from flask samples, continuous analyzers, airborne profiles, and satellite products). We will develop prototype, and evaluate the EnsDA framework using synthetic data generated by the coupled forward model. In the final year of the project, we will apply the EnsDA framework using the fully coupled model constrained by real observations for a test case. This proposal is quite ambitious, but we do not seek support to apply the EnsDA system to real observations on an operational basis. Rather, we will seek operational partners to implement the method in an operational context.

2. Objectives

The central objectives of the proposed research are to ***develop and evaluate a method for providing greatly improved meteorological forcing and trace gas transport in support of NACP, and to apply this method to quantitative estimation of surface carbon fluxes and their uncertainty at regional scales using a dense observing network***. 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 coupled forward model of weather and terrestrial carbon exchange;
2. Develop a generalized EnsDA framework for estimation of control parameters and forward model error;
3. Evaluate the EnsDA system by estimating monthly mean carbon fluxes for one year from synthetic observations using a simplified transport model; and
4. Perform analysis of a test case using the EnsDA framework with the fully coupled model, including assimilation of real meteorological, remote sensing, and atmospheric CO_2 observations.

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

3. Research Plan

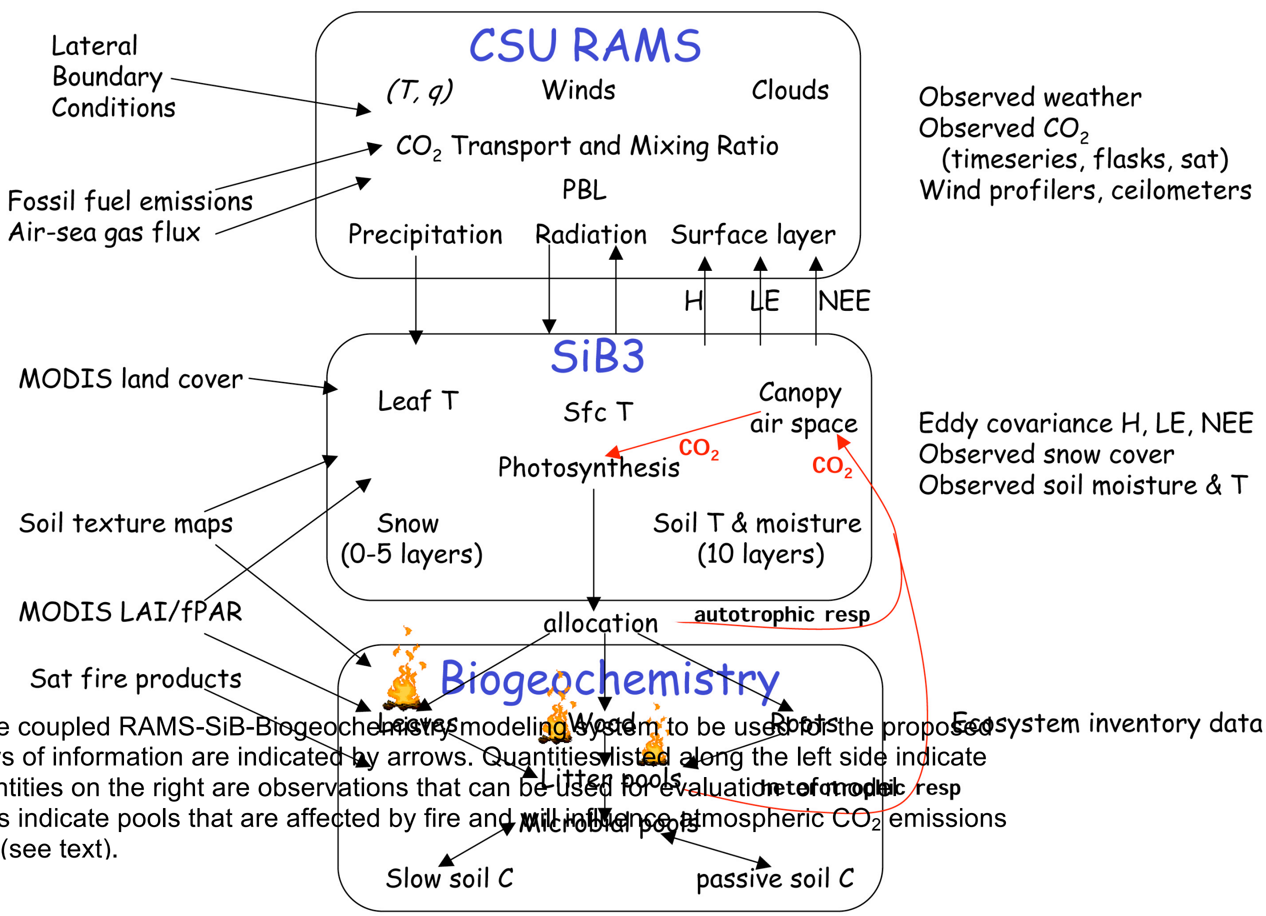
3.1. *Development and testing of the coupled SiB-CASA-RAMS Modeling System*

The Regional Atmospheric Modeling System (RAMS) is a mesoscale meteorological (non-hydrostatic) model and contains time-dependent equations for velocity, non-dimensional pressure perturbation, ice-liquid water potential temperature (Pielke et al, 1992), total water mixing ratio, and cloud microphysics. Vapor mixing ratio and potential temperature are diagnostic. A significant feature of the model is the incorporation of a telescoping nested-grid capability, which enables the simulation of phenomena involving a wide range of spatial scales. A second-order-in-space advection scheme is employed. The turbulence closure scheme of Deardorff (1980) is used, which employs a prognostic sub-grid turbulent kinetic energy. The two-stream radiation scheme developed by Harrington (1997) is used. At regional scales for which individual convective elements (clouds) cannot be resolved, we use convective parameterizations (Grell, 1993; Freitas et al, 2000) that compute precipitation rates, atmospheric heating and moistening, and mass and tracer fluxes (including updraft and downdraft velocities) by unresolved cloud processes. The lowest level above the surface in the RAMS model is the reference level at which atmospheric boundary layer values of temperature, vapor pressure, wind velocity and carbon dioxide partial pressure are calculated. Additionally, the direct and diffuse components of short wave and near infrared radiation incident at the surface are provided from the RAMS radiation scheme.

RAMS has recently been coupled to the Simple Biosphere Model (SiB), 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 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; Sellers et al, 1997; Schaefer et al, 2002). Vegetation type and state are derived from satellite imagery (Sellers et al, 1996b). The model has been updated to include prognostic calculation of temperature, moisture, and trace gases in the canopy air space and has been evaluated for multiyear simulations of a number of eddy covariance sites (Baker et al, 2003; Vidale and Stöckli, 2004). 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 mutilayer snow model based on the Community Land Model (Dai et al, 2003), and the model is now referred to as SiB3. The surface layer, which is between the surface and the reference level of RAMS is incorporated as part of SiB and is based on the scheme of Holtslag and Boville (1993). The input variables provided by RAMS to SiB are updated every minute of simulation time. SiB provides back to RAMS, at the reference level, fluxes of heat, moisture, momentum and carbon dioxide, as well as the upwelling radiation. The coupled SiB-RAMS model has been 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).

put data

Evaluation data



e 4: Structure of the coupled RAMS-SiB-Biogeochemistry modeling system to be used for the proposed
rch. Important flows of information are indicated by arrows. Quantities listed along the left side indicate
inputs, while quantities on the right are observations that can be used for evaluation of model
tions. Fire symbols indicate pools that are affected by fire and will influence atmospheric CO_2 emissions
biomass burning (see text).

Work is ongoing to add a biogeochemistry module to SiB3, based in part on the Carnegie-Ames-Stanford Approach (CASA, Potter et al., 1993). 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 will be 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_2 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. In later data assimilation experiments (see section 3.4 below), we will derive optimal values and uncertainties of these parameters and state variables and the sensitivity of the fluxes to them.

Spin up of the BCM requires use of mean meteorological conditions and GPP for 1000 years with a one-month time step, followed by an additional 100 years with a one-hour timestep. When carbon pools have reached equilibrium, time series of analyzed meteorology and observed vegetation index for the analysis period will be used as boundary conditions to generate hourly carbon fluxes. Such initialization of the analysis with equilibrium conditions precludes study of long-term source and sinks, such as those caused by recovery from disturbance or CO_2 fertilization, but does allow study of climate-driven interannual variability. The spatial distribution of secular sources and sinks will be derived in later stages of the proposed research, using the assimilation methods described below applied to the atmospheric transport model and CO_2 observations. Optimization analysis of the states of relevant carbon pools that could plausibly account for sources and sinks (e.g. live wood pool, coarse woody debris) will identify regions and conditions that could be evaluated with regional information (e.g. Forest Inventory and Analysis, USFS).

Recently, it has been argued that the response of the land surface carbon flux to climate variability is to a large extent the result of climate driven variability in global fires (Langenfelds et al 2002, van der Werf et al 2004). A Co-I on this proposal (Collatz) is part of a NASA funded project aimed at estimating carbon species emissions globally from fires (JR Randerson, PI). Satellite based estimates of burned area and biogeochemical model estimates of fuel loads are being used to estimate monthly CO_2 emissions from fires (van der Werf et al 2004). The team has released monthly fire emissions for the 1997-2001 period (http://www.gps.caltech.edu/~jimr/randerson.html) and will continue to improve and make available to this proposed work emissions estimates through 2007. Relevant aspects of the Randerson et al project will be adopted for SIB-BCM. Emissions will be prescribed from satellite-based estimates of burned area and modeled fuel loads at daily to weekly time steps. Carbon fluxes from fires will include direct emissions caused by fire consumption of biomass and litter pools as well as indirect effects on RH caused by transfers of carbon from killed biomass to litter pools (see Figure 4). The carbon sinks caused by recovery of biomass and litter pools after fire will be simulated as functions of GPP and climate. Satellite vegetation indices should at least in part address the reduction followed by recovery of GPP resulting from destruction of green vegetation and regrowth following fire.

To evaluate the improved model, we will perform a retrospective forward simulation for one year, and compare model predictions of observed quantities at a range of scales to actual observations of meteorological variables, local fluxes of heat, water, and CO_2, and high-frequency variations of atmospheric CO_2 mixing ratio. The period covered by the simulation will likely be 2005, by which time many new CO_2 observing stations and eddy covariance tower sites are expected to be collecting data (Denning et al, 2004). Lateral boundary conditions for meteorological variables will be specified from the NASA Goddard EOS Data Assimilation System (GEOS-DAS) on a 1°x1.25° grid with 55 levels, and will be interpolated on isentropic surfaces to nudge an outer RAMS grid of resolution 100 km that covers all of North America and extends well out over the Atlantic and Pacific Oceans. The outermost three grid columns will be nudged to the global analysis with a three-hour relaxation time. Lateral boundary forcing for CO_2 will be provided from a global simulation with the offline Parameterized Chemical Transport Model (PCTM), which is forced by prescribed fossil fuel emissions, air-sea gas exchange, and hourly net ecosystem exchange simulated by SiB3-CASA driven by MODIS vegetation imagery and GEOS weather analyses (Kawa et al, 2004). In addition, monthly regional fluxes derived by inversion of global observations, to ensure that global mixing ratio fields are optimally consistent with time-mean observations at remote sites. These global analyses will be provided through an ongoing collaboration with scientists at the Goddard Modeling and Assimilation Office through Co-I Jim Collatz. We will run a nested mesoscale grid (Δx=20 km) over the continental USA and adjacent portions of Canada, Mexico, and the oceans, and evaluate the coupled simulation by comparing model quantities with observations within this finer domain. Evaluations will include comparisons of simulated station temperature, humidity, precipitation and winds to local observations; storm events to precipitation radar data; upper air winds to radiosondes and wind profilers; PBL depth to soundings, ceilometers, and profilers; ecosystem fluxes of sensible and latent heat and CO_2 to Ameriflux eddy covariance data; and CO_2 mixing ratio to flask, airborne, and high-frequency observations. Preliminary tests of the coupled modeling system suggest that the proposed simulation experiment would take about 10 days to perform on a single CPU of a state-of-the-art Linux workstation.

It may well be that the first NACP Intensive Field Campaign (http://www.carboncyclescience.gov/nacp-first-intensive-campaign.html) occurs during the time period covered by these simulations, in which case they may be useful for other NACP science. We will make the results of these simulations and comparisons to observations available through the web to interested investigators as well as through the peer-reviewed literature. ***We are not proposing to perform dedicated cloud-resolving nested simulations of the intensively sampled domain***, but such simulations would be enabled by the work we propose here.

3.2. *Ensemble Data Assimilation Framework*

Beginning with the pioneering work of Evensen (1994), many techniques have been proposed, as different variants of EnKF/EnKS (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Keppenne 2000; Mitchell and Houtekamer 2000; Anderson 2001; Bishop et al. 2001; van Leeuwen 2001; Reichle et al. 2002; Whitaker and Hamill 2002; Ott et al. 2004; Tippett et al. 2003; Zupanski 2004; Zupanski and Zupanski 2004). We will develop a general framework for the optimal estimation of model parameters and their uncertainty, which can be used with forward models of arbitrary complexity.

The essence of the EnsDA approach is that an ensemble of sets of systematically perturbed control parameters are generated by the algorithm from an initial forward simulation and calculation of model-data mismatch. A large 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. We will develop optimization software that will read an observation vector and a forward model prediction of the observations. A set of optimally perturbed model parameters will be generated, and an ensemble of forward runs will then be performed, generating an ensemble of new predictions. The EnsDA procedure will then be used to determine optimum values of the model parameters and to estimate the uncertainty associated with each of these parameter values. The only modification to a forward model that will be required to run within the framework will be addition of "observation operators" transforming model variables into quantities that can be directly compared to observations, adding code to write these quantities to output files at the times and locations of real observations, and adding code to read key parameters from input files generated by the optimization algorithm. The general framework for EnsDA will be used in the research proposed herein for improved meteorological analyses, observationally-constrained carbon budgets, and uncertainty estimation in support of NACP, but will be made available for other uses as well.

One of the most critical issues to be resolved when assimilating real observations, is to appropriately estimate and correct model error (Anderson 2001; Hamil et al. 2001; Mitchell and Houtekamer 2000; Reichle et al. 2002; Hansen 2002; Tippett et al. 2003; Zupanski and Zupanski 2004). This problem remains fundamental for future progress in EnsDA methods, ensemble forecasting (Buizza et al. 1999) and predictability Orrell (2003). The forward model error estimation problem will be one of the major foci of this research proposal.

To address the model error issue, we will employ the error estimation methodology recently proposed in Zupanski and Zupanski (2004). We will employ a state augmentation method (e.g., Jazwinski 1970; Gelb 1974), coupled with EnsDA, to estimate a *serially correlated model error*, as well as to estimate *unknown model parameters*. The state augmentation approach has been successfully used to estimate serially correlated model error in variational (Derber 1989; Bennett et al. 1993, 1996; DeMaria, and Jones 1993; Zupanski 1993; Griffith and Nichols 2001; Zupanski 1997; Vidard et al. 2000; D'Andrea and Vautard 2000; Zupanski et al. 2002a,b; 2004) and KF methods (Dee 1995; Dee and da Silva 1998; Martin et al. 1999; Nichols 2003). More recently, the state augmentation approach was used within EnsDA method to estimate model parameters (Anderson 2001; Mitchell and Houtekamer 2000) and model bias (Reichle et al. 2002).

Let us define a new state vector by augmenting the standard model state $\boldsymbol{x}_n$ with model error vector $\boldsymbol{\ddot{O}}_n$. In the case of parameter estimation, the model parameters, rather than vector $\boldsymbol{\ddot{O}}_n$, are used to augment the model state variable. The time evolution of the augmented model state variable, denoted $\boldsymbol{z}_k$, defined in data assimilation cycle *k*, depending also on the model time step *n*, can be described by the following equation:

$$\begin{bmatrix} \boldsymbol{x}_n \\ \text{�}_n \end{bmatrix} = \begin{bmatrix} M\left(\boldsymbol{x}_{n\text{-}1}\right) + \alpha \text{�}_{n\text{-}1} + (1-\alpha)G(\boldsymbol{b}_k) \\ \alpha \text{�}_{n\text{-}1} + (1-\alpha)G(\boldsymbol{b}_k) \end{bmatrix} = F_{0,n} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{b}_k \end{bmatrix} = F_{0,n}\left(\boldsymbol{z}_k\right) \ , \quad n=1,N_{max} \ ; k=1,K_{max} \quad (1)$$

where N_{max} denotes number of model time steps in each data assimilation cycle, and K_{max} is the total number of data assimilation cycles. The forward model (e.g., SiB-CASA-RAMS) is defined as a non-linear operator M, acting upon the model state variable $\boldsymbol{x}$. The augmented control variable $\boldsymbol{z}_k$ includes the following two components: $\boldsymbol{x}_0$ (initial conditions for the forward model) and $\boldsymbol{b}_k$ (model bias, defined as a constant vector for the k-th data assimilation cycle). The empirical constant α measures relative influences of the current ($\boldsymbol{b}_k$) and the previous bias ($\boldsymbol{b}_{k-1}$) on the serially correlated model error $\ddot{\boldsymbol{O}}_n$. The non-linear operator F is a new forward model (or any prediction model depending on initial conditions and model error), where indexes $0,n$ indicate time integration from time step 0 to time step n. The non-linear mapping G, which could depend on the model state $\boldsymbol{x}$, transforms the space of the bias vector ($\boldsymbol{b}$) into the model state vector ($\boldsymbol{x}$) space.

Evaluation

The augmented variable $\boldsymbol{z}_k$ and model F are used in EnKF equations in place of the standard model state $\boldsymbol{x}$ and forecast model M, respectively. For simplicity, we assume that the model error (1) includes only systematic components (the effect of random error is neglected). One can easily include the effects of random noise in model's equations by imposing random perturbations from a prescribed white noise probability density function (PDF).

This study will employ EnsDA methodology and available carbon related observations to provide optimal estimates of the carbon fluxes and uncertainty information of these estimates. In addition, since the methodology will include estimates of model biases as well as the forward model's empirical parameters, the uncertainties of these estimates will also be quantified. An important component of the research to be performed is a quantitative evaluation of the EnsDA results, including the uncertainty estimates. We will employ common evaluation tools used in Kalman filtering experiments, such as the χ^2 – test and probability density function (PDF) of the innovations (observation minus forecast differences).

The χ^2 – innovation test will be calculated (Dee et al, 1995; Menard et al, 2000; Zupanski, 2004). Assume a random variable χ^2 is defined in observation space, normalized by the number of observation, N_{obs}, as:

$$\chi^2 = \frac{1}{N_{obs}}[\boldsymbol{y}_k - H(\boldsymbol{x}_k)]^T[\boldsymbol{HP}_f\boldsymbol{H}^T + \boldsymbol{R}]^{-1}[\boldsymbol{y}_k - H(\boldsymbol{x}_k)] \quad , \tag{2}$$

where $\boldsymbol{y}$ denotes the observation vector, H is the non-linear observation operator, and $\boldsymbol{H}$ is the linearization of it. $\boldsymbol{P}_f$ and $\boldsymbol{R}$ are the forecast and observation error covariance matrices, respectively. For a Gaussian distribution of innovations, and linear observation operator H, the conditional mean of χ^2 should be equal to one.

The PDF of normalized innovations ($\boldsymbol{Y}$) will be calculated employing (e.g., Reichle et al. 2002; Zupanski 2004):

$$\boldsymbol{Y} = \frac{1}{\sqrt{N_{obs}}}[\boldsymbol{HP}_f\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}]^{-\frac{1}{2}}[\boldsymbol{y}_k - H(\boldsymbol{x}_k)] \; . \tag{3}$$

If the non-normalized innovations are assumed Gaussian and white, and the observation operator H is linear, the resulting PDF of normalized innovations ($\boldsymbol{Y}$) has a standard normal distribution $\mathcal{N}(0,1)$.

The innovation statistics (2) and (3) will be used in the proposed study as the basic tests of the algorithm performance. In addition, Root Mean Square (RMS) errors of the optimal solution will be evaluated.

The EnsDA methodology and statistical tests (2) and (3) assume Gaussian distribution of the innovations. It is possible, however, that in applications to the coupled SiB-CASA-RAMS modeling system proposed for this study, the innovation statistics might significantly depart from the Gaussian assumption, rendering the data assimilation results questionable. We will make use of the currently funded NSF research, lead by M. Zupanski, addressing the issue of non-Gaussian assumption within the Maximum Likelihood Ensemble Filter (MLEF) framework (Zupanski 2004). It is anticipated that the research results of the NSF study will provide a guidance for further generalization of the EnsDA methodology, as well as of the verification scores, to be more suitable for non-Gaussian probability distribution functions, which will be of great benefit to the research proposed here.

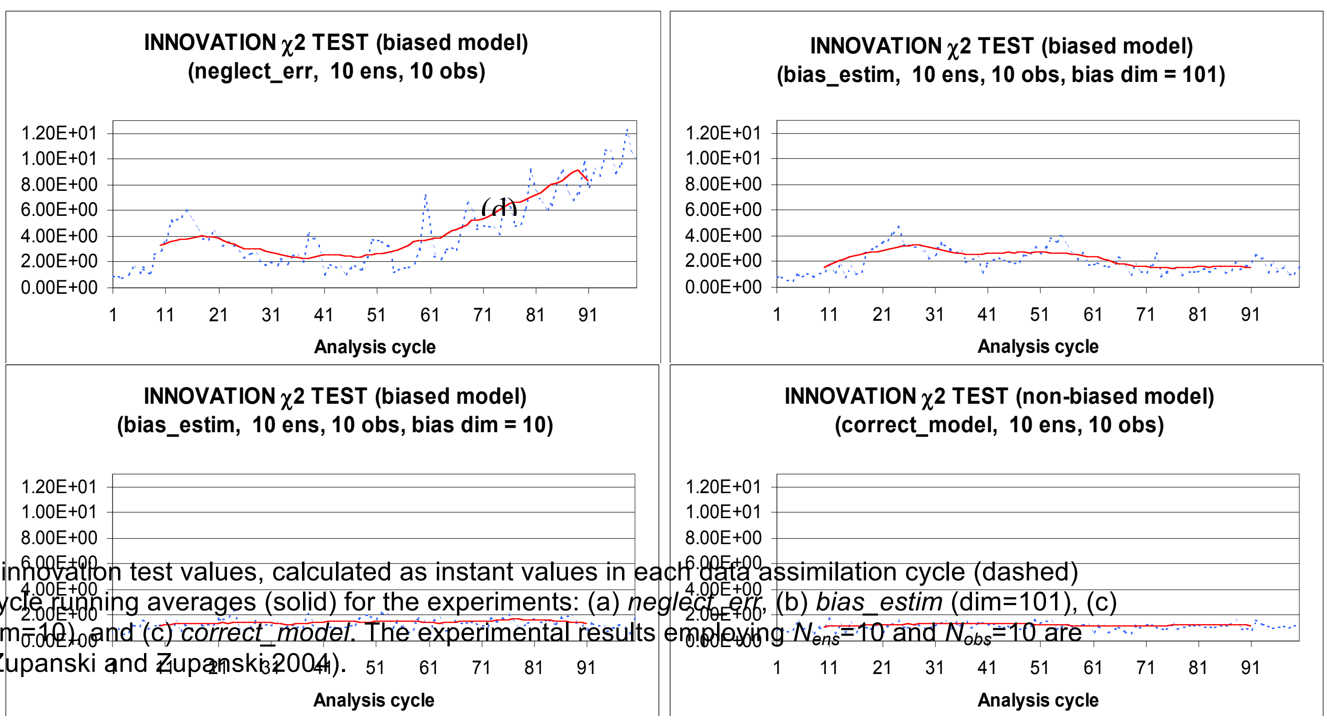
The methodology that will be used for this research is the MLEF approach, proposed by Zupanski (2004), with the model error estimation technique of Zupanski and Zupanski (2004). This is a fully adaptive data assimilation and model error estimation approach, including *optimal estimates* for model state ($\boldsymbol{x}_n$), model error ($\ddot{\boldsymbol{O}}_n$), and model empirical parameters, as well as *uncertainties* of these estimates in terms of the analysis ($\boldsymbol{P}_a$) and forecast ($\boldsymbol{P}_f$) error covariance matrices. The methodology is based on the minimization of a functional (e. g., maximum likelihood approach) via iterative minimization process, which is beneficial in application to non-linear models. An efficient Hessian preconditioning, defined in ensemble subspace, provides very fast minimization convergence (1-3 iterations, Zupanski 2004).

3.3. *Data Assimilation Experiments Using Synthetic Observations*

To test the general EnsDA system in a computationally efficient yet scientifically relevant context, we will perform an assimilation of synthetic CO_2 observations in a simplified tracer transport model. We will first generate 3D fields of atmospheric CO_2 mixing ratio over the North American continent every hour for a year using the forward coupled RAMS-SiB-CASA model as described above (section 3.1). We will then sample the simulated concentration fields at the times, locations, and altitudes of actual NACP observing sites, and use the EnsDA system to estimate monthly mean fluxes and uncertainties. The results of this calculation will be compared to the (known) fluxes from the coupled forward modeling system that produced the simulated concentration field.

Atmospheric transport of CO_2 will be simnulated using an offline Lagrangian Particle Dispersion Model (LPDM) driven from RAMS transport variables (winds, PBL turbulence, cloud mass fluxes) archived every 15 simulated minutes (Ulisaz, 1993, 1994; Uliasz and Pielke, 1991; Uliasz et al, 1996; Uliasz and Denning, 2004). This offline model is extremely fast, yet retains fidelity to the full online transport characteristics of RAMS, facilitating experimentation with a large ensemble of simulations in the EnsDA framework. Surface carbon fluxes will be estimated as monthly mean deviations from a "background" (a priori) field that we will produce using biome-specific light-response curves and temperature-dependent ecosystem respiration fluxes (similar to the method of Gerbig et al, 2003b). Empirical parameters in the background flux model (light and temperature sensitivities, baseline rates) and monthly time-mean fluxes and associated uncertainties will be estimated by the EnsDA procedure. This procedure allows isolation of the slowly-varying error in the prior fluxes from the high-frequency variations that can be reasonably modeled, and mimics the temporal error covariance structure in the real world. Lateral boundary fluxes of CO_2 will be treated as a "soft" constraint in this calculation, with corrections to the "background" global field and uncertainties estimated b ythe assimilation procedure (Uliasz and Denning, 2004).

These experiments will be critical for evaluating the overall performance of the EnsDA algorithm. Similarly as in Zupanski and Zupanski (2004), the experimental results of model bias estimation, as well as estimation of model empirical parameters will be examined. The innovation statistics (2) and (3), as well as RMS errors with respect to the truth, will be used as evaluation tools. Since the observation error $\boldsymbol{R}$ is perfectly known in the experiments with synthetic observations, the test results of the innovation statistics (2) and (3) will measure the correctness of the calculated forecast error covariance matrix ($\boldsymbol{P}_f$). A significant departure from the expected test results would indicate incorrect $\boldsymbol{P}_f$, and filter divergence. To illustrate the usefulness of the innovation statistics, we show in Figure 5 an example of χ^2 – innovation test of the proposed EnsDA methodology, in application to Korteweg-de Vries-Burgers (KdVB) numerical model. The following EnSDA experiments are presented: (a) *neglect_err*, neglecting model bias, (b) *bias_estim* (dim=101), estimation of model bias of dimension 101 (c) *bias_estim* (dim=10), estimation of model bias of dimension 10 and (c) *correct_model*, employing a model without an error (non-biased model). The experimental results employing 10 ensembles and 10 synthetic observations (with known errors), in each data assimilation cycle, over 100 cycles, are shown. Fig. 5a indicates filter divergence for the experiment *neglect_err*! Bias estimation process is clearly beneficial (Figs. 5b,c), with the values of χ^2 much closer to the results of the *correct_model* (Fig. 5d) than to the experiment *neglect_err* (Fig. 5a). These results indicate that it is critical to appropriately take model bias into account in order to provide a reliable uncertainty estimates and prevent filter divergence. The results of the preliminary study of Zupanski and Zupanski (2004) also indicated the potential of the EnsDA methodology to correctly estimate



igure 5: χ^2 – innovation test values, calculated as instant values in each data assimilation cycle (dashed) nd as a 10-cycle running averages (solid) for the experiments: (a) *neglect_err*, (b) *bias_estim* (dim=101), (c) *ias_estim* (dim=10), and (c) *correct_model*. The experimental results employing N_{ens}=10 and N_{obs}=10 are hown (from Zupanski and Zupanski 2004).

unknown model parameters (figures not shown, more results can be found in the paper).

3.4. *Prototype Data Assimilation Using the Coupled Model and Real Observations*

In the final year of the proposed research, we will apply the EnsDA framework to estimation of model parameters, uncertainty, and forward model error in the fully coupled RAMS-SiB-CASA model using all available real observations. Choice of parameters to optimize will be driven by (1) a need to estimate parameters that are poorly known yet important for obtaining accurate net carbon exchange; (2) restriction to quantities that exert strong influence on observable quantities; (3) availability of observational; and (4) computational efficiency. We plan to estimate magnitudes and uncertainties in the following parameters, as well as quantify model error:

- Buoyancy and TKE dependence of eddy diffusivity at the simulated PBL top;
- Lateral boundary CO_2 mixing ratios;
- Initial ecosystem carbon pool sizes (including spatial variations);
- Combustion efficiency in the fire module, and dependence on moisture status;
- Ecosystem drought stress dependence on soil moisture;
- Biome-dependent sensitivity of GPP to direct and diffuse light;
- Temperature sensitivity of autotrophic respiration and decomposition; and
- The residual monthly-mean flux of CO_2 at each model grid cell.

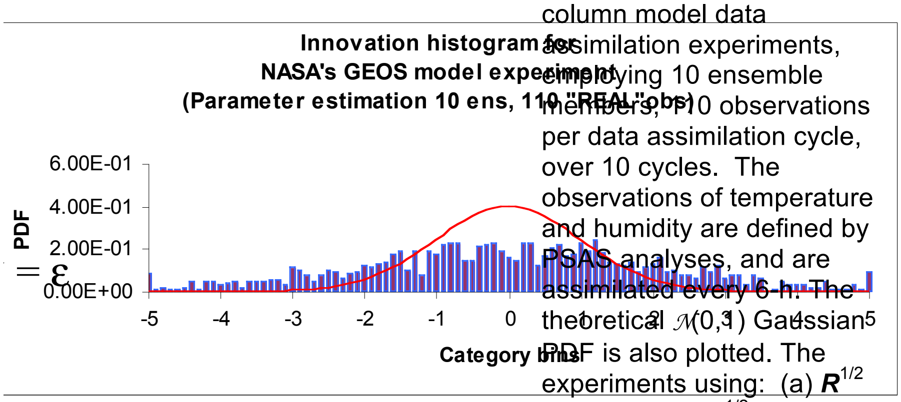
Observational constraints against which these parameters will be optimized will include:

- MODIS land cover and vegetation state (LAI, fPAR);
- fire occurrence and areal extent;
- eddy covariance data measurements of sensible and latent heat flux and net ecosystem exchange of CO_2;
- observations of PBL depth from radiosondes, wind profilers, RADARs and SODARs; and
- atmospheric CO_2 measured on flask samples, airborne profiles, and continuous analyzers on tall towers, mountains, and eddy covariance (“virtual tall”) towers, as well as satellite products (based on AIRS and OCO, if available).

It is possible, of course, that some of the observations will not have realistic observation errors assigned, and sometimes these errors are unknown. Assuming preliminary tests with simulated observations confirm that the calculated $\boldsymbol{P}_f$ is reliable, the evaluation tools (2) and (3) can be used to test the specification of observation error covariance ($\boldsymbol{R}$). An example of such test is given in Figure 6. The experimental results, in terms of PDF of normalized innovations (3) are presented. The same EnsDA algorithm as in Zupanski and Zupanski (2004) is employed, but in application to a column version of the NASA’s Goddard Earth Observing System (GEOS) global forecast model. Global analyses (PSAS, Cohn et al. 1998) are used to define observations (temperature and humidity) every 6-h, during 10 data assimilation cycles. As the figure indicates, the PDF of innovations shows sensitivity to different choices of observation errors (1/2 ε vs. ε). These results demonstrate that it is possible to use the innovation statistics to obtain a reliable estimate of ***R***. We will test this hypothesis in the proposed study, in application to carbon data assimilation.

/2= 1/2 ε

It is anticipated that the proposed research will benefit from related data assimilation research, led by D. Zupanski, under a currently funded NOAA/NESDIS research project. The main objective of the NOAA/NESDIS project is to reduce the risk of the future GOES-R satellite mission. A data assimilation algorithm employing RAMS atmospheric model and the EnsDA methodology will be developed during the three-year course of the GOES-R research project (2003-2005). The atmospheric EnsDA algorithm, developed under the NESDIS study, will be coupled with the carbon data assimilation algorithm of the proposed study. This will result in a possibility to examine the EnsDA methodology in application to the models of higher complexity then ever before (the coupled RAMS-SiB-CASA model). We will use this opportunity to perform a demonstration study, employing the coupled model and simulated atmospheric and carbon observations, to assess the potential benefits of the EnsDA methodology in applications to state-of-the-art coupled atmospheric-hydrology-biosphere models in carbon cycle studies.



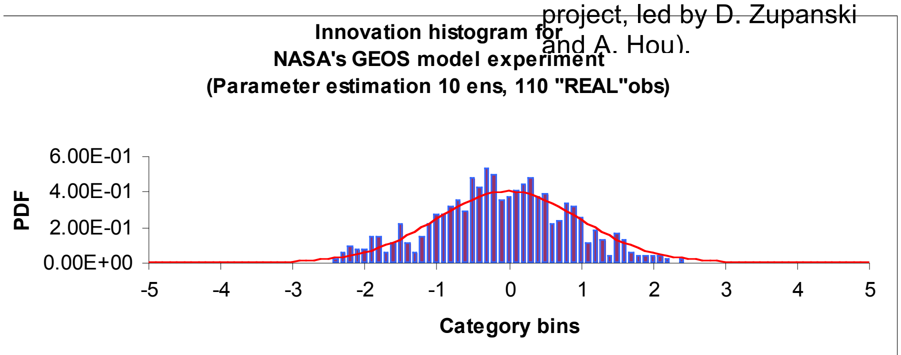


Figure 6: Histogram of the PDF of the innovations, calculated from the GEOS column model data assimilation experiments, employing 10 ensemble members, 110 observations per data assimilation cycle, over 10 cycles. The observations of temperature and humidity are defined by PSAS analyses, and are assimilated every 6-h. The theoretical $\mathcal{N}(0,1)$ Gaussian PDF is also plotted. The experiments using: (a) $\boldsymbol{R}^{1/2}$ = ½ ε and (b) $\boldsymbol{R}^{1/2}$ = ε are presented. (Unpublished results of a NASA research project, led by D. Zupanski and A. Hou).

The computational requirements of this calculation are much more stringent than the assimilation of synthetic observations described above, because a large ensemble (~ 100 members) of fully coupled year-long model simulations (Δx=20 km, Δt ~ 60 seconds) will be run with perturbed parameter values and lateral boundary conditions. Preliminary calculations indicate that the full ensemble would take about 1000 days on a single-CPU workstation. We are therefore requesting funding for a 20-CPU Linux computer cluster and RAID storage system to accommodate these very ambitious computational requirements, and estimate that this will reduce the time required for EnsDA analysis of one year of data in the coupled modeling system to less than two months.

The EnsDA system with the coupled model could be extended to include assimilation of all available meteorological observations (thousands of surface stations, hundreds of daily radiosondes, aircraft data, weather RADARs, and wall-to-wall coverage by satellite meteorological products). Such extension could provide weather reanalysis at arbitrarily high resolution in space and time for driving ecosystem and hydrologic models and completely self-consistent atmospheric transport fields to drive inverse calculations of carbon fluxes, but ***we are not proposing to perform this analysis***. The work proposed here will develop and test a method by which such analyses could be performed, but the actual production of these analyses is beyond the scope of the proposed work. We will however, work closely with operational agencies (e.g., NOAA laboratories) to disseminate our research findings, and will actively seek partners to implement customized coupled meteorological data assimilation for carbon cycle research.

4. Schedule of Work

The proposed research will be performed over a period of three years. During this time the coupled RAMS-SiB-CASA mdoel will be further developed and tested. The EnsDA methodology, applicable to carbon data assimilation problems, will be developed and applied to estimate carbon fluxes and uncertainties of these estimates, taking into account model errors. The EnsDA algorithm (already developed and tested in applications to the KdVB model and GEOS column model) will be used as a starting point. The transport and particle dispersion models will be used in cycled data assimilation experiments using synthetic observations generated by the forward coupled model to produce analyses over a one-year period that can be rigorously evaluated against known fluxes. In the final stage of the proposed research, the fully coupled RAMS-SiB-CASA model will be employed in a prototype EnsDA to estimate magnitudes and uncertainties of time-mean fluxes, control parameters, lateral boundary conditions, and model error 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 World Wide Web.

The schedule of specific tasks to be performed is as follows

Year 1: Algorithm development

- Continue development of coupled RAMS-SiB-CASA model, including implementation of carbon allocation, biogeochemistry, and fire modules;
- Perform short test experiments with RAMS-SiB-CASA over limited domains and compare in detail to local and regional observations of meteorology, carbon flux, and atmospheric CO_2;
- Develop the basic EnsDA algorithm for application to mesoscale carbon data assimilation employing transport and particle dispersion models.
- Include model bias and parameter estimation.
- Prepare a web-page documenting the work progress.
- Present the results at the meetings, scientific conferences and prepare manuscripts for scientific journals. Prepare research reports.

Year 2: Data assimilation experiments with synthetic observations

- Perform a 1-year forward simulation with the coupled RAMS-SiB-CASA model, and evaluate against available observations of weather, ecosystem fluxes, and CO_2 mixing ratio;
- Specify an NACP atmospheric carbon observing system and generate synthetic observations from the forward simulation;
- Run the receptor-oriented offline transport model (LPDM) for each observation, driven by the archived RAMS transport fields;
- Perform data assimilation experiments employing model generated observations.
- Evaluate data assimilation results in terms innovation statistics and RMS errors.
- Evaluate data assimilation results in terms innovation statistics and other verification tools.
- Perform cycled data assimilation experiments to obtain optimal estimates for carbon fluxes, transport model errors, unknown model parameters, and uncertainties of these estimates in terms of analysis error covariance matrix ($\boldsymbol{P}_a$).
- Archive results and make available through a dedicated web site;
- Continue updating the web-page with the work progress and research results.
- Present research results at the meetings, conferences, scientific papers, etc. Prepare research reports.

Year 3: Prototype EnsDA Analysis Using a Complex Coupled Model

- Develop observation operators to relate coupled model predictions to real observations;
- Obtain one year's observational data (meteorology, satellite imagery, eddy covariance data, and in-situ CO_2 mixing ratios);
- Develop an observational error covariance specification in collaboration with data providers;
- Adapt coupled RAMS-SiB-CASA model to include input of control parameters and output of observational operators
- Perform demonstration ensemble data assimilation experiments using the complex EnsDA algorithm and real observations;
- Archive results and make available through a dedicated web site;
- Continue updating the web-page with the work progress and research results.
- Present research results at the meetings, conferences, scientific papers, etc.
- Prepare final report.

5. References cited

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev*., **129**, 2884–2903.

Andres, R. J., G. Marland, I. Fung, and E. Matthews (1996), A 1º x 1º distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950– 1990, *Global Biogeochem. Cycles*, **10**, 419– 429.

Baker, I.T., A.S. Denning, N. Hanan, L. Prihodko, P.-L. Vidale, K. Davis and P. Bakwin, 2001: 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.

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.

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.

Barker, M., 2004: Model error estimation using advanced data assimilation systems. Submitted to *Mon. Wea. Rev*. (also available at ftp://ftp.cira.colostate.edu/Zupanski/manuscripts/RAMDAS_paper_MWR.pdf).

Bender, M. et al, 2002. *A large-scale CO2 observing plan: In situ oceans and atmosphere (LSCOP)*. NOAA OAR Special Report, 201 pp. Available online: http://www.ogp.noaa.gov/mpe/gcc/co2/observingplan/toc.htm

Bennett, A. F., L. M. Leslie, C. R. Gagelberg, and P. E. Powers, 1993: Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Wea. Rev*., **121**, 1714–1729.

Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys*., **60**, 165–178.

Bishop, C. H., B. J. Etherton, and S. Majumjar, 2001: Adaptive sampling with the ensemble Transform Kalman filter. Part 1: Theoretical aspects. *Mon. Wea. Rev*., **129**, 420–436.

Buizza, R., M. Miller and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. Jour. Roy Met. Soc*.,**125**, 2885-2908.

CarboEurope Integrated Project, 2003. *Assessment of the European Terrestrial Carbon Balance*. Sixth Framework Programme (Global Change and Ecosystems), Contract No. 505572, European Union, 276 pp.

Christi, M. J. and G. L. Stephens, 2004. Retrieving profiles of atmospheric CO 2 in clear sky and in the presence of thin cloud using spectroscopy from the near and thermal infrared: A preliminary case study, *J. Geophys. Res*., **109**, D04316, doi:10.1029/2003JD004058.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*., **75**, 257–288.

Cohn, S.E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical–space statistical analysis system. *Mon. Wea. Rev*., **126**, 2913-2926.

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_4 plants. *Aust. J. Plant Physiol.*, **19**, 519-538.

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.

D'Andrea, F., and R. Vautard, 2000: Reducing systematic errors by empirically correcting model errors. *Tellus*., **52A**, 21–41.

Dee, D., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev*., **123**, 1128–1145.

Dee, D., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc*., **124**, 269-295.

DeMaria, M., and R. W. Jones, 1993: Optimization of a hurricane track forecast model with the adjoint model equations. *Mon. Wea. Rev*., **121**, 1730–1745.

Derber, J., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev*., **117**, 2437–2466.

Denning, A.S., I.Y. Fung and D.A. Randall, 1995: Latitudinal gradient of atmospheric CO_2 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_2 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_2 in a general circulation model. Part 2: Spatial and temporal variations of atmospheric CO_2. *Tellus*,

48B, 543-567.
Denning, A.S. et al, 1999: Three-dimensional transport and concentration of SF_2: A model intercomparison study (TransCom 2). *Tellus,* **51B**, 266-297.
Denning, A.S., T. Takahashi and P. Friedlingstein, 1999. Can a strong atmospheric CO_2 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_2 over a Wisconsin forest using a coupled Ecosystem-Atmosphere Model.. *Global Change Biology*, **9**, 1241-1250.
Denning, A. S., et al, 2004. *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.
Doney, S.C. et al, 2004: *Ocean Carbon and Climate Change (OCCC): An Implementation Strategy for U. S. Ocean Carbon Cycle Science*, UCAR, Boulder, CO, 104pp.
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 CO2 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.
Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res*., **99**, (C5),. 10143-10162.
Farquhar, G. D., von Caemmerer, S. and Berry, J. A., 1980. A biochemical model of photosynthetic CO_2 assimilation in C_3 plants. *Planta*, **149**, 78-90.
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.
Gelb, A., 1974: *Applied Optimal Estimation*. The MIT Press, 374 pp.
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_2 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_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.
Ghil, M. et al, 1981: *Applications of estimation theory to numerical weather prediction. Dynamic Meteorology: Data Assimilation Methods*., L. Bengtsson, M. Ghil, and E. Källén, Eds., Springer-Verlag, 139–224.
Grell, G. 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev*, **121**, 764-787.
Griffith, A. K., and N. K. Nichols, 2001: Adjoint techniques in data assimilation for treating systematic model error*, J. of Flow, Turbulence and Combustion*, **65**, 469-488.
Gurney, K. R., et al. (2002), Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, *Nature*, **415**, 626– 630.
Gurney, K. R., et al, 2003. TransCom3 CO_2 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.
Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter/3D-variational analysis scheme. *Mon. Wea. Rev*., **128**, 2905–2919.
Hansen, J.A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev*., **130**, 2373–2391.
Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev*., **126**, 796–811.
Hurwitz, Michael D. et al., 2004. Transport of carbon dioxide in the presence of storm systems over a northern Wisconsin forest. *Journal of the Atmospheric Sciences,* **61**: 607-618 .
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 CO2 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 and Z. Zhu, 2004. Global CO_2 transport simulations using meteorological data from the NASA data assimilation system, *J. Geophys. Res*., Submitted.

Langenfelds, R. L. et al., 2002. Interannual growth rate variations of atmospheric CO_2 and its $\delta^{13}C$, H_2, CH_4, 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 CO2 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_2 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_2 inversions using high frequency data, *Tellus*, **55B**, 512–521, doi:10.1034/j.1600-0560.2003.0029.

Martin, M. J., N. K. Nichols, and M. J. Bell, 1999: *Treatment of Systematic Errors in Sequential Data Assimilation*, UK Meteorological Office, Ocean Applications Division, Tech. Note No. 21, 45 pp.

Menard, R., S.E. Cohn, L.-P. Chang, and P.M. Lyster, 2000: Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation. Kalman filter. *Mon Wea. Rev*., **128**, 2654-2671.

Mitchell, H. L, and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev*., **128**, 416–433.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc*., **122**, 73–119.

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,* in press.

Nichols, N. K., 2003: Treating Model Error in 3-D and 4-D Data Assimilation, Data Assimilation for the Earth System, *Proceedings of the NATO Advanced Study Institute*, Acquafredda, Maratea, Italy, May, 2002 (Eds R. Swinbank, V. Shutyaev, W.A. Lahoz), Kluwer Academic.

Orrell, D. 2003: Model error and predictability over different timescales in the Lorenz '96 systems. *J. Atmos. Sci*, **60**, 2219–2228.

Ott, E. et al, 2004: A local ensemble Kalman filter for atmospheric data assimilation. Posted http://arXiv.org/abs/physics/0203058. Submitted to *Mon. Wea. Rev*.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc*., **74**, 49–64.

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_2 data, *J. Geophys. Res*., **107**, 4385, doi:10.1029/2001JD000857.

Pielke, R. A. et al, 1992. A comprehensive meteorological modeling system - RAMS. *Meteor. Atmos. Phys*., **49**, 69-91.

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_2, $\delta^{13}C$, and O_2/N_2 observations. *Tellus*, **51B**, 213-232.

Rayner, P.J. and D.M. O'Brien, 2001. The utility of remotely sensed CO2 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_2 from space, *J. Geophys. Res*., **107**, 4557, doi:10.1029/2001JD000618.

Reichle, R. H., D. B. McLaughlin, D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev*. **130**, 103–114.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann, 2003. Time dependent atmospheric CO_2 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_2 fluxes. *Global Biogeochemical Cycles,* **16**, 1102, doi:10.1029/2002GB001928.

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_2 fluxes, *Eos Trans*. **CG62A-05**.

Sellers, P.J. et al,, 1996*a*: A Revised Land-Surface Parameterization (SiB2) for Atmospheric GCMs. Part 1: Model

formulation. *J. Clim.*, **9**, 676-705.
Sellers, P. J. et al,, 1996*b*: 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 Program.* A Report by the Climate Change Science Program, Washington, DC: US Global Change Research Program. Available on-line http://www.climatescience.gov/Library/stratplan2003/final/default.htm
Suits, N.S., A.S. Denning, J.A. Berry, C.J. Still, J. Kaduk and J.T. Randerson, 2004. Seasonal and spatial variations in carbon isotopic ratios of plant biomass, terrestrial CO_2 fluxes and atmospheric CO_2. *Global Biogeochemical Cycles*, in press.
Takahashi, T. et al, 1999. *Net sea-air CO_2 flux over the global oceans: An improved estimate based on the sea-air pCO_2 difference*, paper presented at 2nd International CO_2 in the Oceans Symposium, Cent. for Global and Environ. Res., Natl. Inst. For Environ. Stud., Tsukuba, Japan.
Tippett, M., et al, 2003: Ensemble square-root filters. *Mon. Wea. Rev*., **131**, 1485–1490.
Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc*., **74**, 2317–2330.
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.
Uliasz, M. and A. S. Denning, 2003. Deriving mesoscale surface fluxes of trace gases from concentration data. Submitted to *Journal of Applied Meteorology*.
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 Leeuwen, P. J., 2001: An ensemble smoother with error estimates. *Mon. Wea. Rev*., **129**, 709–728.
Vidard, P.A., E. Blayo, F.-X. Le Dimet and A. Piacentini, 2000: 4D variational data analysis with imperfect model. *Flow, Turbulence and Combustion*, **65**, 489-504.
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.
Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*., **130**, 1913–1924.
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 D. and M. Zupanski, 2004: Model error estimation employing ensemble data assimilation approach. Submitted to *Mon. Wea. Rev*. (available at ftp://ftp.cira.colostate.edu/Zupanski/manuscripts/MLEF_model_err.revised.pdf).
Zupanski D., M. Zupanski, E. Rogers, D. F. Parrish and G. J. DiMego, 2002a: Fine resolution 4DVAR data assimilation for the Great Plains tornado outbreak of May 3rd 1999. *Wea. Forecasting*, **17**, 506-525.
Zupanski, D., M. Zupanski, T. Vukicevic, T. Vonder Haar, D. S. Ojima, W.-S. Wu and D. Zupanski, M., 2004: Maximum likelihood ensemble filter: Theoretical aspects. Submitted to *Mon. Wea. Rev*. (available at ftp://ftp.cira.colostate.edu/Zupanski/manuscripts/MLEF_MWR.pdf).
Zupanski M., D. Zupanski, D. Parrish, E. Rogers and G. J. DiMego, 2002b: Four-dimensional variational data assimilation for the blizzard of 2000. *Mon. Wea. Rev*. **130**, 1967-1988.

A. Scott Denning, PI

Dusanka Zupanski, Co-PI

Colorado State University

Mesoscale Carbon Data Assimilation for NACP

Management Plan

Senior Personnel:

P.I. 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.

Co-PI 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 precipipation 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.

Other Personnel:

Jim Collatz is an accomplished terrestrial ecophysiologist and biogeochemist. He will lead the development, evaluation, and implementation of new algorithms and model code for phenology, carbon allocation, autotrophic respiration, litterfall, microbial processes, and decomposition in SiB, based on his previous experience as a developer of the CASA ecosystem model. He will also be responsible for improved algorithms for the use of MODIS phenology, vegetation state, and fire products in the modeling system.

Milija Zupanski is an internationally-known expert on meteorological data assimilation. He spent 12 years at NCEP, where he was the principal developer of a 4DVAR data assimilation system for NCEP's Eta regional operational forecast model, and after coming to CSU developed the data assimilation algorithm for CSU RAMS. He will work with the other scientists on the project to extend the EnsDA system to estimate and correct model transport error.

Marek Uliasz has over 20 years of experience in regional transport modeling, Lagrangian particle dispersion calculations, and source/sink estimation from atmospheric trace gas observations. He will be responsible for offline tracer transport simulations, the treatment of lateral boundary conditions and their uncertainty, and for testing the EnsDA system with synthetic data.

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

John Kleist has nearly 30 years of experience as a scientific and systems programmer. He will be responsible for building and maintaining the computer cluster and RAID storage system, and related system software. He will also work with other project staff and students to coordinate computing jobs and assist in software engineering of all kinds in support of the research. He will be responsible for optimization and parallelization of the model codes, for implementing the EnsDA system in the parallel computing environment, for load management on the cluster, and for data visualization, archival, and backup.

Connie Uliasz will be responsible for technical writing tasks, and for logistical duties associated with communication and collaboration. She will coordinate the dissemination of model analyses on the World Wide Web (see below).

One of the graduate research assistants will work primarily on source/sink estimation from synthetic atmospheric observations using the Lagrangian Particle Dispersion Model driven from RAMS output. The other will focus on parameter and uncertainty 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!

Dissemination of Research Results

Preliminary research results will be presented each year by the five people (D. Zupanski, M. Uliasz, Baker and the two GRAs) funded under this project to attend the Annual Fall AGU meeting in San Francisco. Denning and Collatz will also be attending this meeting to present their research results.

Analyses and model simulations of weather, carbon and energy fluxes, and atmospheric CO_2 will be made available through the world wide web for other investigators to use in support of NACP-related research.

For publications in peer-reviewed scientific journals, funding has been requested in Years Two and Three.