PARAMETER ESTIMATION IN THE COMMUNITY LAND MODEL USING THE DATA ASSIMILATION RESEARCH TESTBED

Andrew M. Fox[1,2], William J. Sacks[2], David J. P. Moore[3], Tim Hoar[2], Steve Berukoff[1], David S. Schimel[1]

[1] National Ecological Observatory Network; [2] National Center for Atmospheric Research; [3] University of Arizona

Background

Interactions between the climate system and vegetation exhibit complex feedbacks. Climate dynamics control many aspects of ecological function, whilst changes in vegetation influence carbon, water and energy budgets directly affecting local and global climate. The National Ecological Observatory Network (NEON) is a continental scale facility that will collect biogeochemical and biogeophysical data from 60 sites in the US over 30 years.

Estimates of these fluxes at regional and continental scales are required to diagnose, understand and predict the response of the global water and carbon cycles to a changing climate. To perform both spatial extrapolation from NEON sites and temporal forecasting we are developing a model-data fusion framework in which NEON data can be combined with the Community Land Model. Our goal is to produce optimal solutions for model states, fluxes and parameter values, with their associated uncertainties, at regional to continental scales.

Community Land Model (CLM)

The Community Land Model (CLM) is used as the land component in the Community Earth System Model (CESM). It simulates terrestrial ecosystem processes including the cycling of energy, water, carbon and nitrogen. CLM is driven by a limited set of climate variables, which may come from site observations, reanalysis or a coupled atmospheric model, while the sensitivity of ecosystem processes to climate is controlled by the initial states and parameter sets of the model.

An initial step has been to evaluate the performance of CLM-CN at a number of existing flux tower sites with long data records including Harvard Forest, Howland Forest and Niwot Ridge. Looping available site climate observations we spin up the model for 2000 years under pre-industrial conditions, then use a 150 year long transient run with increasing CO_2 concentrations and nitrogen deposition up to the present day. Typically, we find energy fluxes to be more accurately simulated than carbon fluxes (Figure 1).

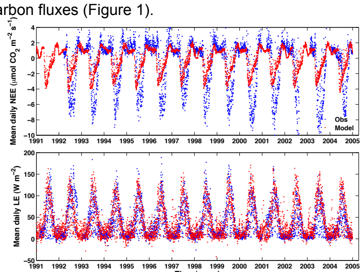

Figure 1. Observed (blue) and modeled (red) carbon and water fluxes at Harvard Forest FLUXNET site

Data assimilation in carbon cycle science

Data assimilation is a general term for methods that systematically combine information from observations with information from a model to achieve an understanding of the system that is more accurate than the observations or the model independently. The data assimilation approach adopted by carbon cycle scientists draws on tools developed in meteorology and applied mathematics to support numerical weather prediction. The goal of carbon data assimilation is often parameter, rather than state, estimation (as is usually the case in NWP). The reason for this is that parameter estimates in carbon cycle models give insight into process-level responses to environmental variation, *e.g.*, the temperature sensitivity of respiration or the photosynthetic response to humidity. In addition, carbon modeling typically probes coupled systems with very different time constants (minutes to decades or longer) that must be considered simultaneously, whereas effects of "slow" geophysical processes typically appear in NWP models as initial or boundary conditions. While reanalysis of atmospheric and oceanic data can reveal the role of slowly varying (typically oceanic) processes, diagnosis of the role of slower processes is a key challenge in carbon modeling. Our approach has been to use ensemble filter techniques, approximate Monte Carlo solutions to the DA problem that have grown rapidly in popularity since their first description in the 1990s (Evensen 1994). By using careful software engineering, it is possible to develop a state-of-the-art ensemble filtering system that is mostly independent of the geophysical model and observations being assimilated. Such a system is the Data Assimilation Research Testbed (DART), a community facility for ensemble DA developed and maintained at the National Center for Atmospheric Research (NCAR) which provides a number of enhancements to basic filtering algorithms.

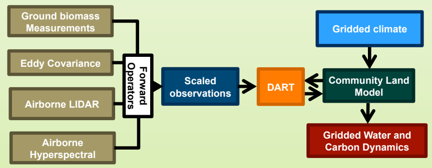

Figure 2. CLM/NEON data assimilation scheme

Parameter sensitivity analysis

Prior to estimating parameters through data assimilation we are undertaking parameter sensitivity analyses to identify parameters to optimize. At first we are assessing parameters contained in the plant functional type (PFT) physiology parameter namelist file. Here we show results from perturbing four parameters: (i) leaf carbon to nitrogen ratio (*leafcn*); (ii) fraction of leaf N in rubisco (*flnr*); (iii) specific leaf area at top of canopy (*slatop*); and (iv) allocation ratio between fine root carbon and leaf carbon (*froot_leaf*). Each parameter was perturbed individually over a range of 0.3 to 1.7 of the default value at 0.1 increments to give 15 ensemble members, the eighth member being equivalent to the default value, and the model then run forward 100 years, initialized with restart files created from a transient run. Each model ensemble member was compared with flux tower observations over the first 13 years and perturbing these parameters individually over this range does not result in modeled carbon fluxes in better agreement with observations. However, it is clear that the model responds to the parameter perturbation at a number of different time scales and ***the long-term effects of the parameter perturbation can differ greatly from the short-term effects*** (Figure 3). This is, in part, because the parameter perturbation alters the quasi-equilibrium state of CLM reached after spin-up, and the short-term carbon dynamics reflect the resultant re-adjustment of the whole model state, rather than processes controlled by the parameters directly. This has important implications for data assimilation, high-lighting the ***difficulty in relying on NEE measurements alone and the need for additional data constraints***. We will also have to address how to ***best initialize a data assimilation run that relies on a previously spun-up model state***.

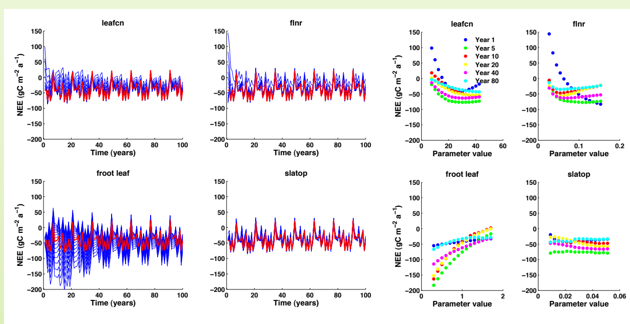

Figure 3. Annual NEE calculated using default parameter values (red) and 14 parameter value perturbations (blue) and scatter plots of annual NEE against parameter value at six time intervals

Perfect Model Experiment

We are currently setting up a perfect model, or observing system simulation, experiment (PME). Starting from a 'spun up' global CLM-CN case for year 2000 we advance a 40-member ensemble of the model globally at 1.9 x 2.5 degree resolution for 120 days at 6hrly time steps. Each instance of CLM is coupled with an individual from an ensemble of data atmospheres, each an equally plausible output from a previous DA exercise with the CESM atmosphere model. This causes a divergence in CLM states across the ensemble. This spread of CLM states provides the initial conditions for the PME. A single instance of CLM from the ensemble is then assigned to be the 'truth' and run forward. Daily 'observations' of leaf carbon (*leafc*) are collected at sites (grid cells) across the globe by adding Gaussian noise to the actual *leafc* state at these locations. These 'observations' are then assimilated by DART as the whole 40-member ensemble is run forward. Below is an illustration of preliminary results showing priors and posteriors of leaf carbon in a model grid cell at 60°W, 4°S for six days when synthetic observations were assimilated using DART.

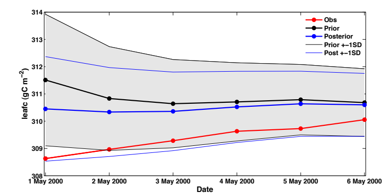

Figure 4. Prior and Posterior distributions of leaf carbon in a single grid cell for six days of assimilation

Future Work

We will continue undertaking a series of PMEs to test the ability of the ensemble filter, as implemented in DART, to update a number of CLM states. We will use a suite of synthetic observations of carbon and water fluxes and pools available at different temporal frequencies based on those that will be made at NEON sites in the future and which are available as remote sensing products.

We will then extend the DART state vector to include CLM parameters and test how we can us this ensemble approach in parameter optimization. Will we investigate how this approach works with parameters which control processes operating at very different timescales

In tandem we will investigate how to assimilate data into a spun-up model when the true initial state of the system at the time when observations are available is unknown and the impacts of parameter optimization on model spin up.

Contact Information: afox@neoninc.org

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc. This material is based in part upon work supported by the National Science Foundation under the following grants: EF-1029808, EF-1138160, EF-1150319 and DBI-0752017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.