Using data assimilation to reduce uncertainties in carbon stocks and fluxes in an Earth System Model
Andrew M. Fox1 and Tim Hoar2
1 National Ecological Observatory Network; 2 National Center for Atmospheric Research

Background
Earth system models that simulate the carbon cycle across a coupled land, ocean and atmosphere are important tools to attribute and predict changes of atmospheric greenhouse gases at regional and global scales and the impacts on climate. The land model component of these schemes is the source of much prediction uncertainty that results from uncertainties associated with representation of model process, parameter values, initial conditions and atmospheric forcing. All of these sources of potential uncertainties are important, but we suggest that to enable accurate forecasting it is essential to fully characterize initial conditions at the start of the prediction period. This is particularly critical when there is large deviation between spin-up model states and observed carbon stocks due to specific disturbance histories.

We are combining eddy covariance flux data and additional measurements from the continental-scale National Ecological Observatory Network platform and other monitoring networks (FLUXNET, Ameriflux, ICOS, LTER etc.) along with satellite remote sensing products with a land surface model using data assimilation methodologies to better characterize current carbon stocks and fluxes across North America.

Data Assimilation
Data assimilation (DA) is the systematic combination of data and models, taking into account the uncertainties in both. The process model provides an analytical framework for data interpretation, synthesis, interpolation and extrapolation. An ensemble filter combines a prior model ensemble, an observation and its likelihood to compute an updated ensemble estimate and corresponding increments to the prior ensemble. In the idealized example below:

1. A prior member ensemble (blue asterisks) is observed to time (blue green lines).
2. A forward operator (h) is applied to obtain estimates of observations (green ticks on upper axes).
3. All observations are then combined with the prior ensemble estimate to obtain an ensemble that matches the observed ensemble (blue deck).
4. The increments are then reinserted into each state vector component independently to generate updates for the state vectors.
5. The filter advances these states to time (red) when the next observation is available.

Idealized ensemble Kalman filter

Clim-DART
Processes simulated by the Community Land Model
We have successfully coupled the well established, open source Community Land Model (CLM-CN), which simulates biogeophysical, hydrological and biogeochemical processes, with the Data Assimilation Research Testbed (DART), an advanced community facility for ensemble data assimilation developed and maintained at the National Center for Atmospheric Research.

Assimilating NEE and LAI observations
Having spun-up CLM-CN using observed climate forcing at the Ameriflux site at Niwot Ridge, CO, we then generated a 64 member ensemble forward run from 1 January 2000 using an ensemble of data atmospheres (DATMs) generated in a previous DA exercise with the Community Atmosphere Model. Each DATM is different and this generates spread in the land model ensemble.

In an observation system simulation experiment (OSSE) we treat one ensemble member as ‘truth’ and sample with appropriate noise NEE every half hour and LAI every 5 days from 1 June 2004. We then investigated the impacts of assimilating these synthetic observations in three experiments: (1) assimilating just LAI; (2) assimilating just NEE; (3) assimilating both LAI and NEE.

This “sawtooth” plot shows LAI simulated by all 64 ensemble members in exp. 1. The increments (updates) calculated by the filter for the prior ensemble towards the observations result in a reduction in uncertainty (spread) around the truth.

Comparing exp. 1 and 3 against a free run (no assimilation), we see using both LAI and NEE observations gives a better result, both in terms of ensemble mean truth accuracy and tight confidence intervals that encompass the truth.

12 months of NEE observations were assimilated in exp. 2 and 3 (17,520 half hourly observations). For clarity, mean daily values for the ensemble means are plotted here and it is clear there is little impact of the assimilation on NEE fluxes themselves.

Zooming in the final 3 days of the assimilation period, assimilation has only very minor impacts on ensemble means and spread. It is apparent that the variability in DATMs at this time scale generates very little variance across the ensemble at night.

Updating Unobserved Variables
Over 4 years there is a small increase in ensemble spread in the free run (blue) for LAI and total ecosystem carbon stocks. Assimilation constrains LAI uncertainty to < 20% and total ecosystem C to c.50% of free run values. LAI uncertainty increases again over the 3 year forecast period, returning to values similar to the free run. However, total ecosystem C uncertainty does not increase rapidly and is still constrained to <40% of free run values after 3 years of forecast.