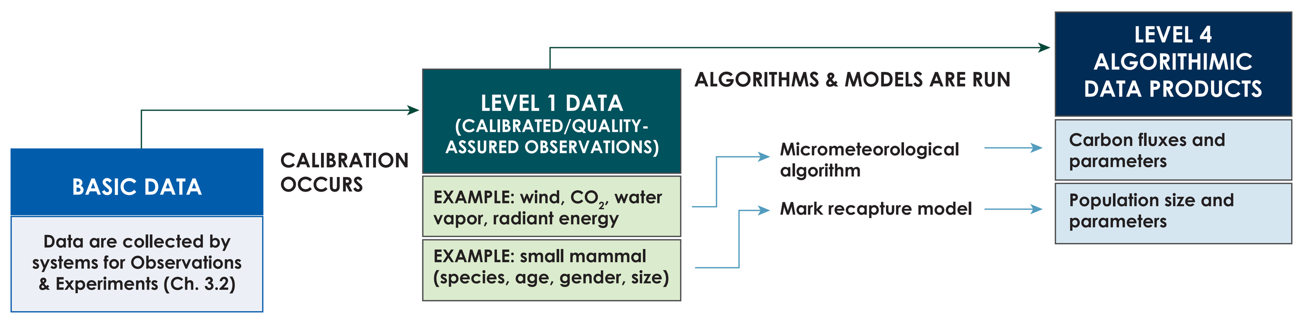
The impacts of uncertainty in ecological observations on a data assimilation system

Andrew M. Fox[1], Jeffery Taylor[1], Tim Hoar[2]

[1] National Ecological Observatory Network; [2] National Center for Atmospheric Research

Background

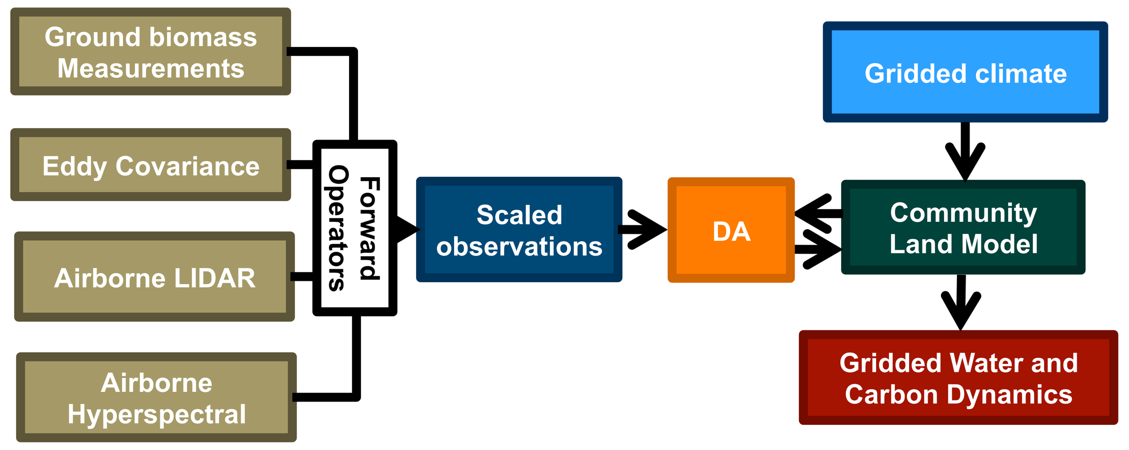
We are vigorously quantifying uncertainty in all NEON data products and are investigating the propagation of uncertainty from basic observations when generating high-level data products, such as continental-scale, gridded maps of carbon and water fluxes. One approach we are using to integrate many observational data streams into high-level data products is to use data assimilation. In this study we investigate how different assessments of uncertainty in eddy covariance flux measurements interact to affect our ability to quantify and reduce uncertainty in high-level data products.



Basic flow of NEON data products

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. If done well model states become more consistent with observations (and hopefully the 'truth') and forecasts become more accurate as initial conditions are improved.

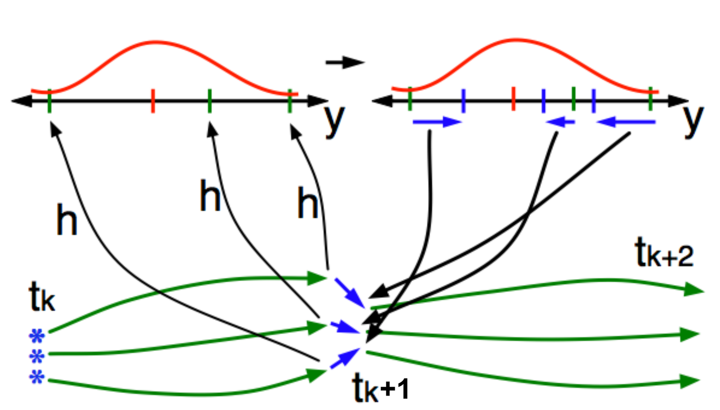


A data assimilation system for NEON observations

Ensemble Filters

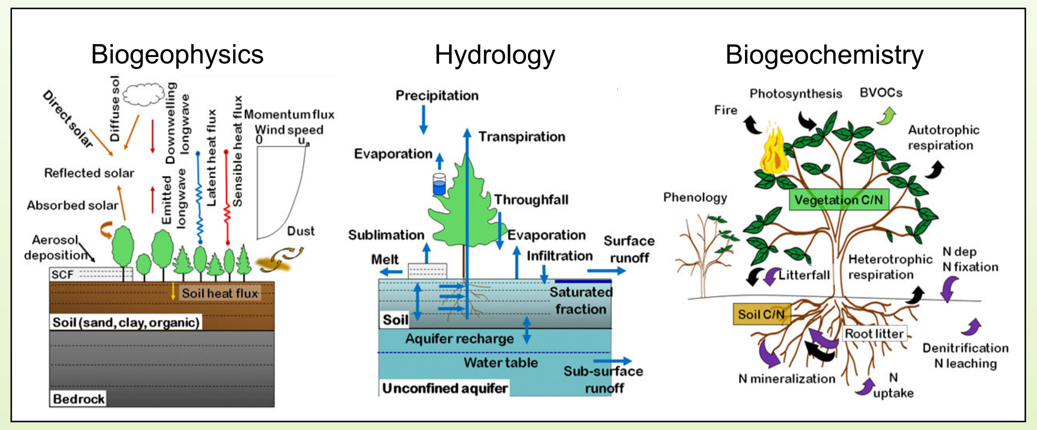
An ensemble filter combines a prior model ensemble, an observation and its likelihood to compute an updated ensemble estimate and corresponding increments to the prior ensemble. In the idealized example below:

I. A three member ensemble (blue asterisks) is advanced to time t_{k+1} by a model (green lines).
II. A forward operator (*h*) is applied to obtain estimates of observations (green ticks on upper axes, y).
III. An observations (red tick) and its likelihood (red curve) are combined with the prior ensemble estimate to obtain increments and an updated ensemble (blue ticks).
IV. The increments are then regressed onto each state vector component independently to generate updates in the state vector.
V. The model then advances these states to time t_{k+2} when the next observation is available.



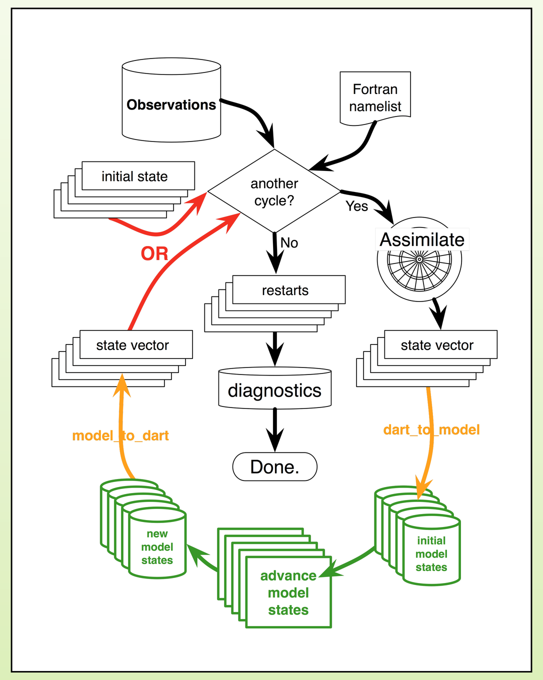
Idealized operation of an ensemble Kalman filter

CLM-DART



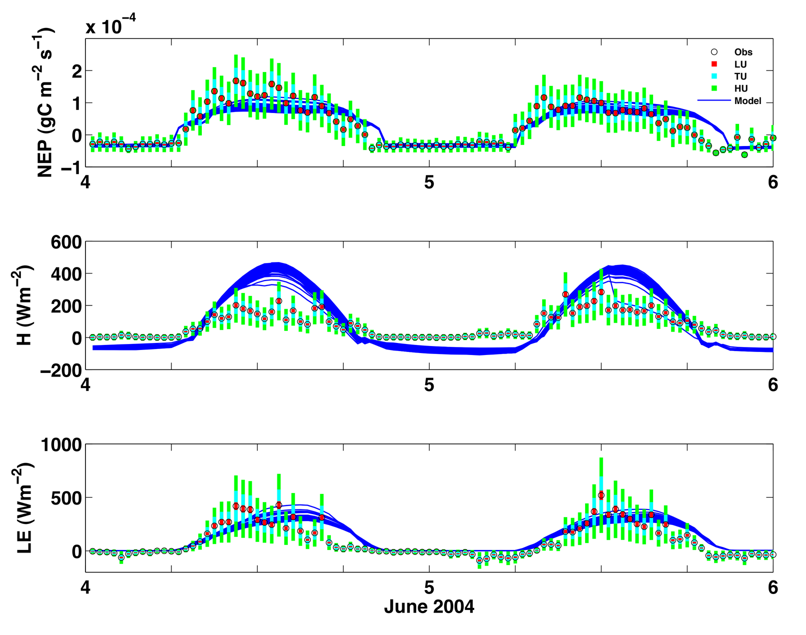
Processes simulated by the Community Land Model

We have successfully coupled the well established, open source Community Land Model (CLM) (Oleson et al., 2010), with the Data Assimilation Research Testbed (DART) (Anderson et al., 2009), an advanced system for ensemble data assimilation.



DART-model coupling

Flux measurement uncertainties

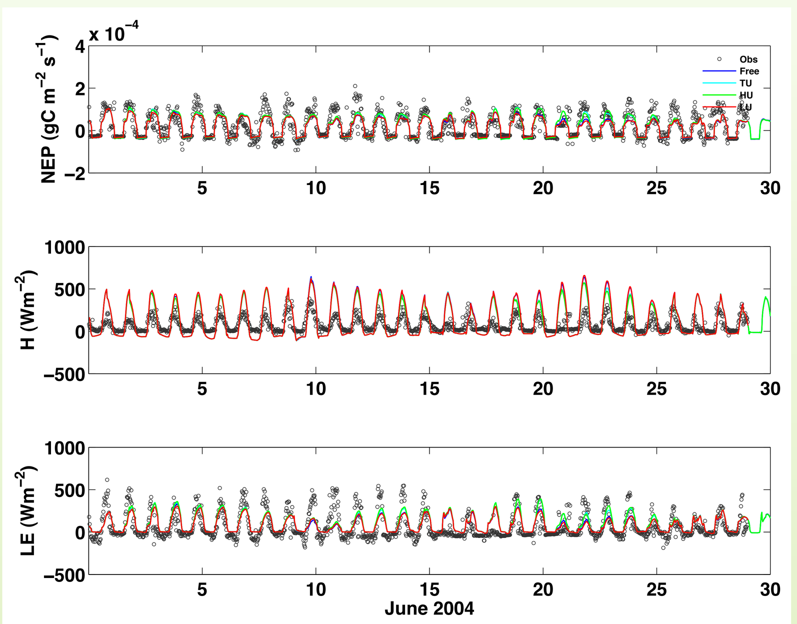


Low, true and high uncertainties on two days of ecosystem flux measurements

We calculated three levels of uncertainty on fluxes from Niwot Ridge for June 2004 that we could then assimilate. "True" uncertainty is taken from Hollinger and Richardson (2005)

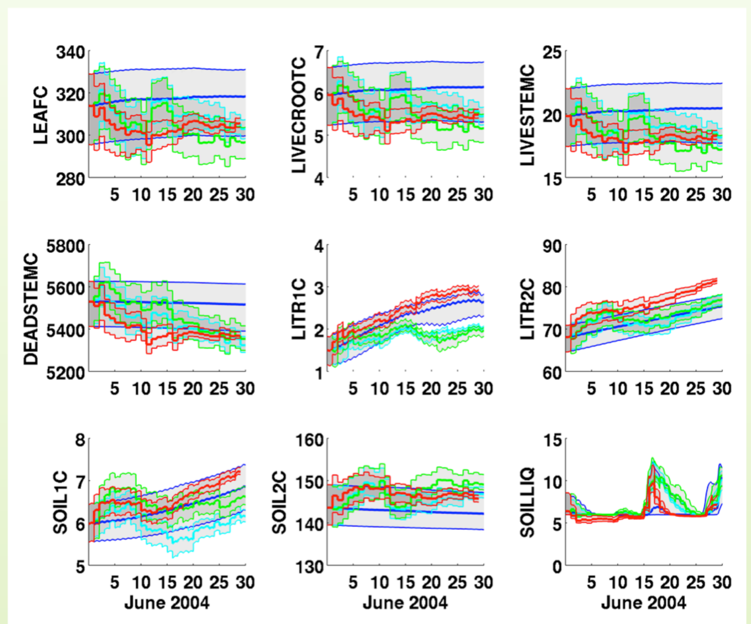
Level	Flux	Uncertainty
Low (LU)	H (sensible heat)	2.5 + 0.055 |H|
	LE (latent heat)	2.5 + 0.08 |LE|
	NEP (net ecosystem productivity)	0.5 + 0.025 NEP (NEP > 0)
		0.5 + 0.1 NEP (NEP < 0)
True (TU)	H	10 + 0.22 |H|
	LE	10 + 0.32 |LE|
	NEP	2 + 0.1 NEP (NEP > 0)
		2 + 0.4 NEP (NEP < 0)
High (HU)	H	20 + 0.44 |H|
	LE	20 + 0.64 |LE|
	NEP	4 + 0.2 NEP (NEP > 0)
		4 + 0.8 NEP (NEP < 0)

Observed and Unobserved model variables



Observations and mean model ensemble estimates at Niwot Ridge

Surprisingly, it can be seen that the ensemble assimilating the high uncertainty (HU) observations is adjusted the most by DART.

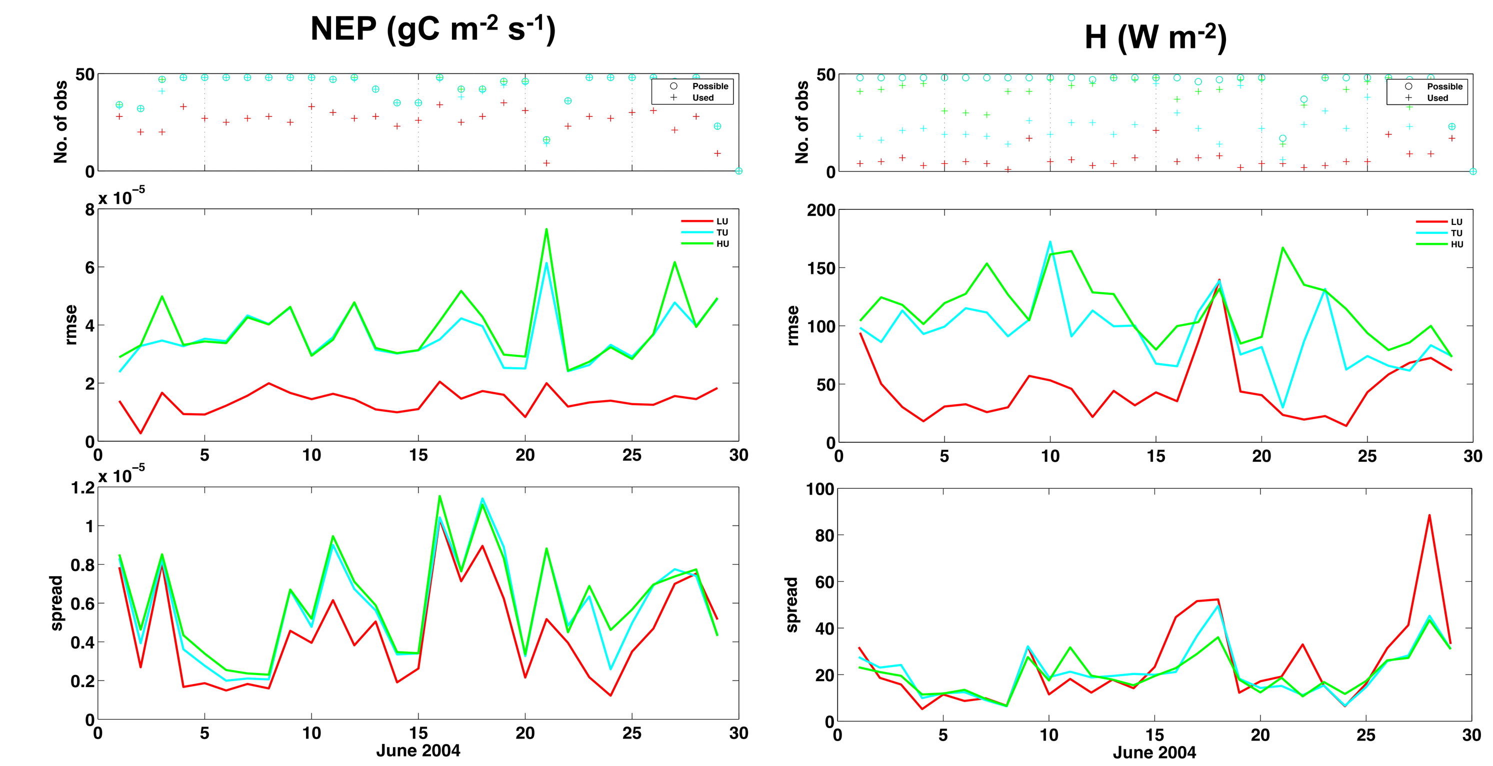


DART can also adjust many unobserved carbon and water state variables

By including them in the DART state vector the error covariance matrix is able to update many unobserved variables based on correlation with observed variables.

Diagnostics

So why are we seeing this surprising result? Filter theory would suggest that the observations with the highest uncertainties would have the least impact on the ensemble.



Diagnostic plots showing the number of observations assimilated successfully each day, ensemble error and ensemble spread

Whether an observation is used depends on its quality flag (only 'good' observations are used) and distance of the observation from the ensemble mean. Here we used a distance threshold of three ensemble standard deviations. Many of the low uncertainty observations have likelihoods too far from the ensemble mean and are excluded. Filter performance can be seen to be highly sensitive to the number of observations used.

Future work

We continue to test the performance of CLM-DART using Ameriflux eddy covariance flux data and other ecological observations available at a number of future NEON core sites. We will also investigate using DART's sophisticated tools for adaptive inflation and localization to improve filter performance.
This is an ongoing project and many theoretical and technical questions remain as to how to construct an operational system for generating high-level NEON data products.

These include:

I. How to create initial ensemble spread – how large should it be?
II. How to maintain ensemble spread – is climate forcing variability the best approach?
III. What do we do about carbon/water balance – it's lost at the moment and balance checks are removed?
IV. What are the most informative observations to use – and can we develop appropriate forward operators to link them with CLM state?
V. How can we best use an ensemble DA approach for parameter estimation – we can augment DART state vector with CLM parameters, but which ones?

References

Hollinger D. Y. and A. D. Richardson, 2005. Uncertainty in eddy covariance flux measurements and its application to physiological models. *Tree Physiol* **25**:873-885
Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009. The Data Assimilation Research Testbed: A Community Facility. *Bulletin of the American Meteorological Society*, **90**:1283-1296.
Oleson, K.W., et al., 2010: Technical Description of version 4.0 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-478+STR, National Center for Atmospheric Research, Boulder, CO, 257 pp.

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 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.