QUANTIFYING UNCERTAINTY IN PROJECTIONS OF CONTINENTAL FLUXES OF CARBON AND WATER USING THE NEON PLATFORM

Andy Fox[1], Tim Hoar[2]

1. National Ecological Observatory Network
2. Institute for Mathematics Applied to Geosciences (iMAGE), National Center for Atmospheric Research

Why are we interested in this at NEON?

	LOCAL (SITES)	REGIONAL (AOP)	NATIONAL (LUAP)	NATIONAL (DATA + MODELS)
Biomass/Productivity/Metabolism	✓	✓	✓	✓
Soil Structure/Physics	✓	✓	✓	✓
Bioclimate/Energy Balance	✓	✓	✓	✓
Hydrology/Ecohydrology	✓	✓	✓	✓
Habitat/Landscape Structure	✓	✓	✓	✓
Biodiversity/Invasive Species/Biogeography	✓	✓		✓
Population Dynamics/Demography	✓		✓	✓
Land Use Land Cover		✓	✓	✓
Biogeochemistry	✓	✓		
Microbial Diversity/Function	✓			✓
Atmospheric/Air Quality	✓		✓	
Phenology	✓			
Infectious Diseases/Vectors	✓			
Ecological Stoichiometry	✓			

Its an extensive network, but…

Still requires massive extrapolation

- Area of the USA = 9,826,675 km²

- Area of flux tower footprint ~ 1 km²
- 60 tower locations at NEON and ~60 towers in other networks with equivalent QA/QC
- **0.0012%** of the USA is sampled

- Area of NEON airborne platform ~400 km²
- **0.244%** of the USA is observed

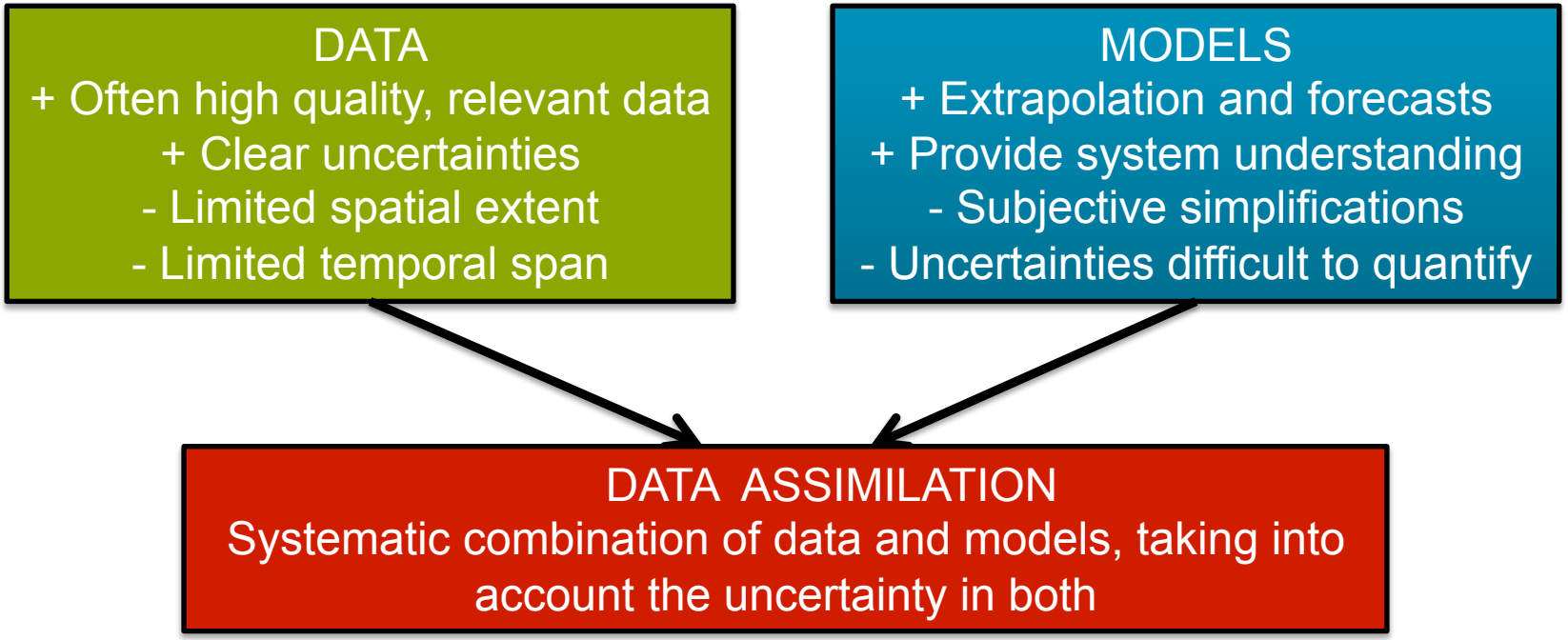
- We're going to have to do something clever to generate those data products…

DATA ASSIMILATION

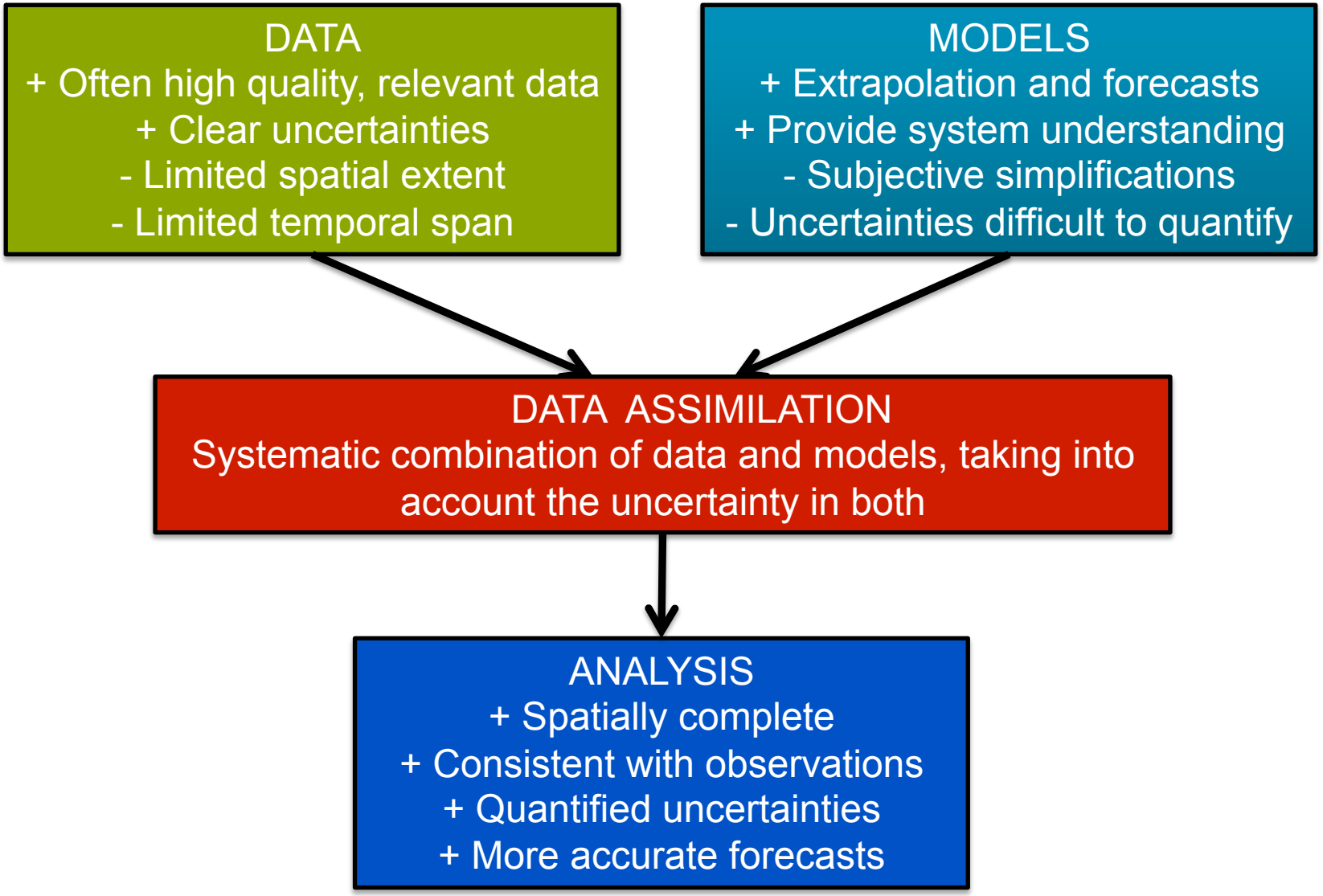
What is data assimilation?

DATA ASSIMILATION
Systematic combination of data and models, taking into account the uncertainty in both

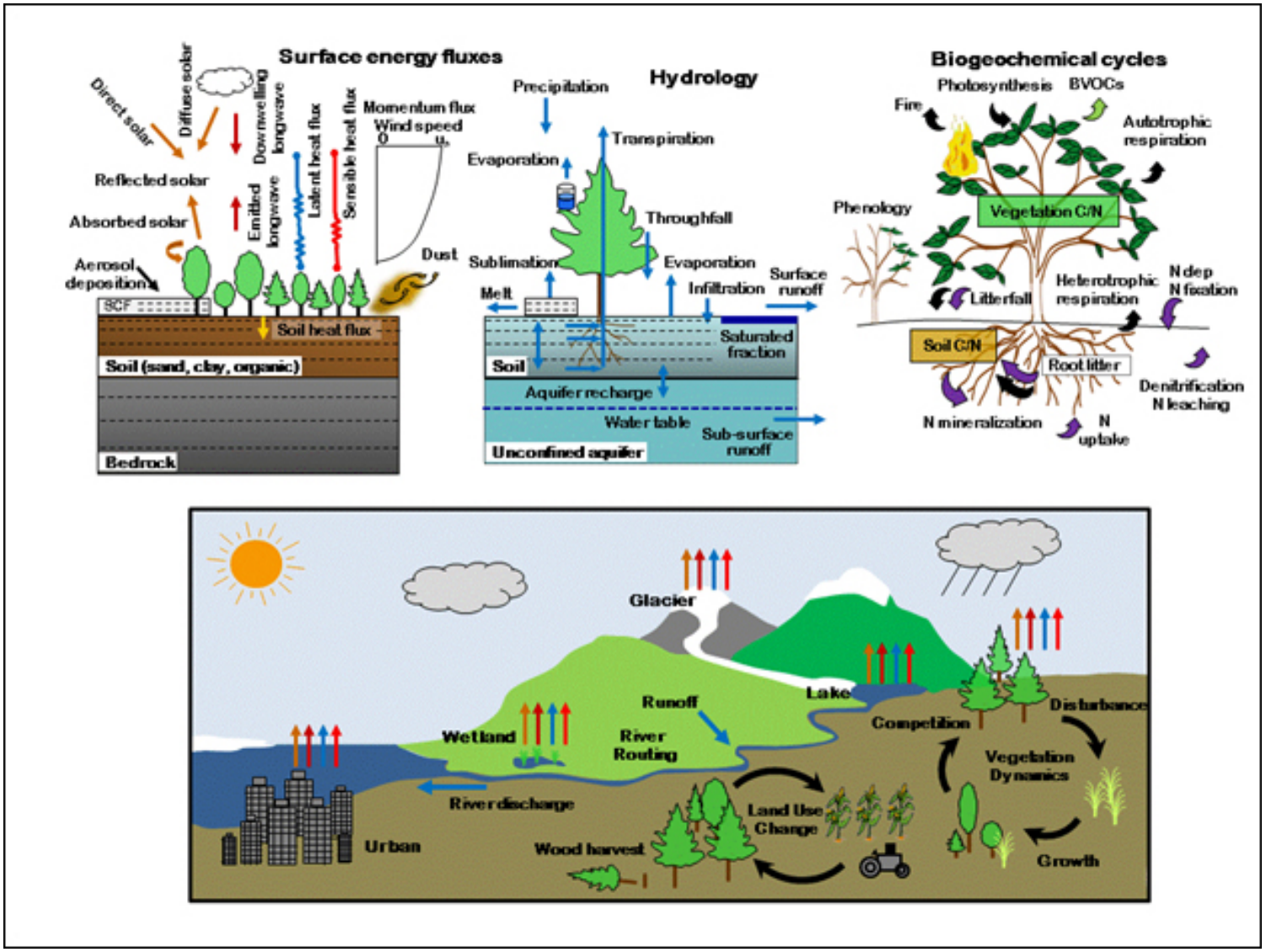
What is data assimilation?



What is data assimilation?

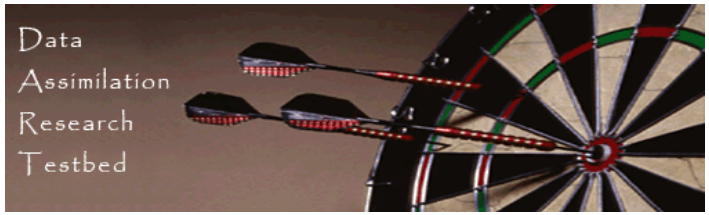


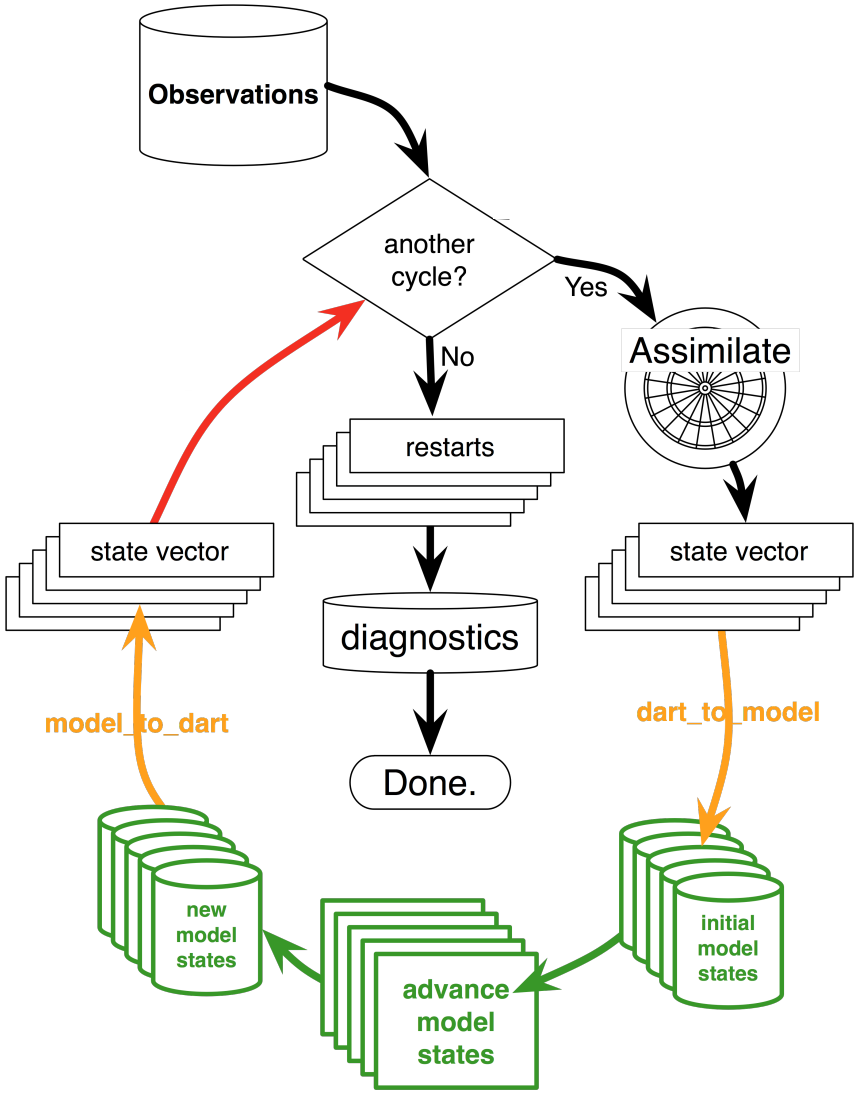
The Community Land Model (CLM)



Data Assimilation Research Testbed (DART)

- DART is a community facility for ensemble DA
- Uses a variety of flavors of filters
  - Ensemble Adjustment Kalman Filter
- Many enhancements to basic filtering algorithms
  - Adaptive inflation
  - Localization
- Uses new multi-instance capability within CESM





Doing this is a BIG computing task

420,000 core-hours on Yellowstone to continue this work

ENSEMBLE FILTERS

Ensemble filter for large geophysical models

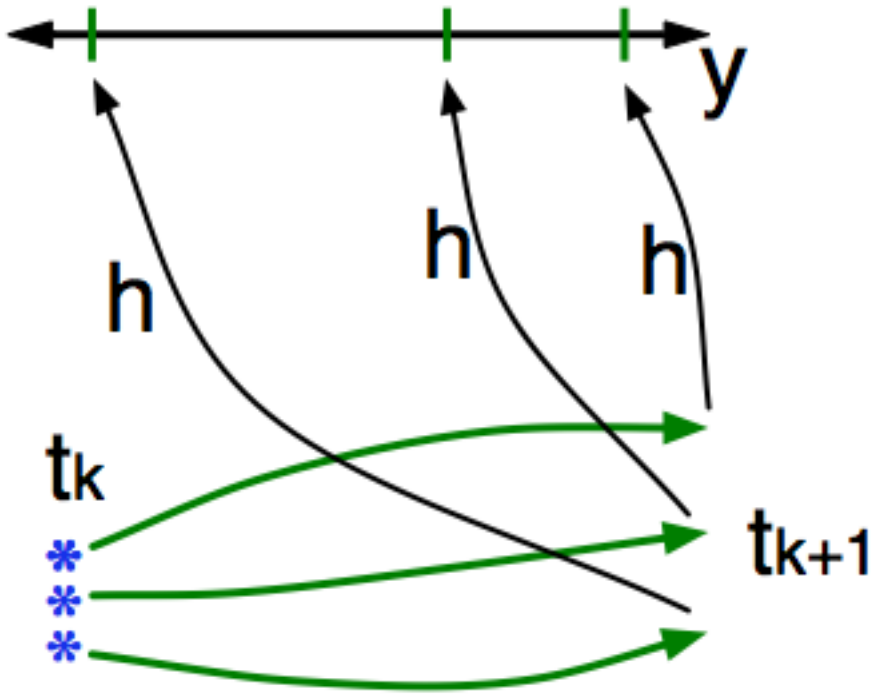
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

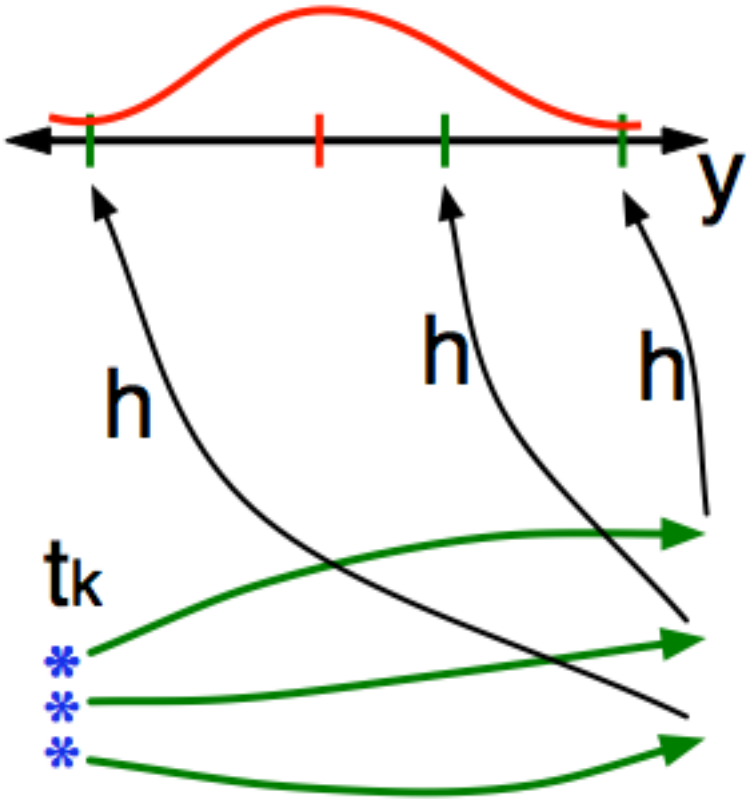
Ensemble filter for large geophysical models

2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.



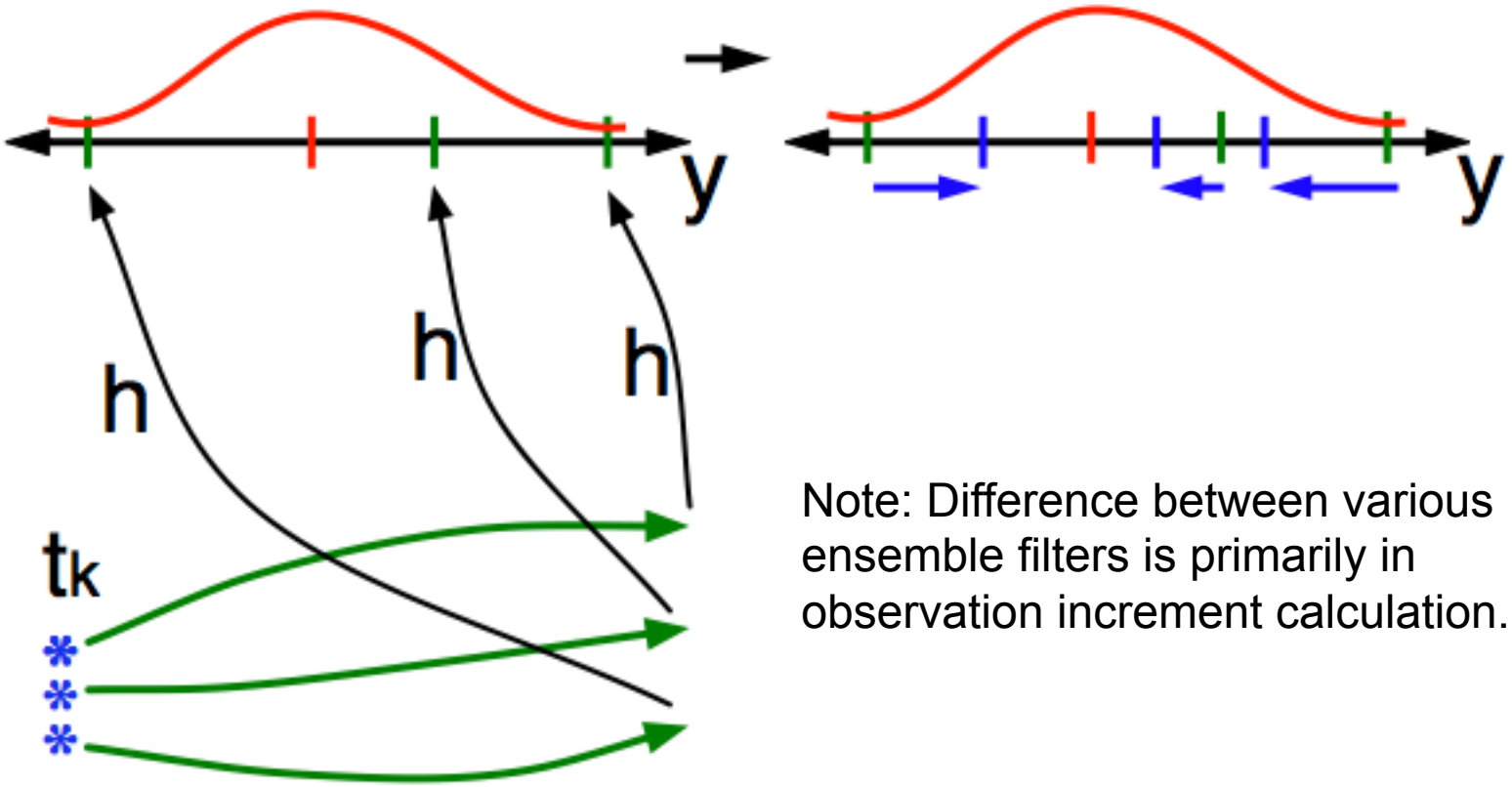
Ensemble filter for large geophysical models

3. Get observed value and observational error distribution from observing system.



Ensemble filter for large geophysical models

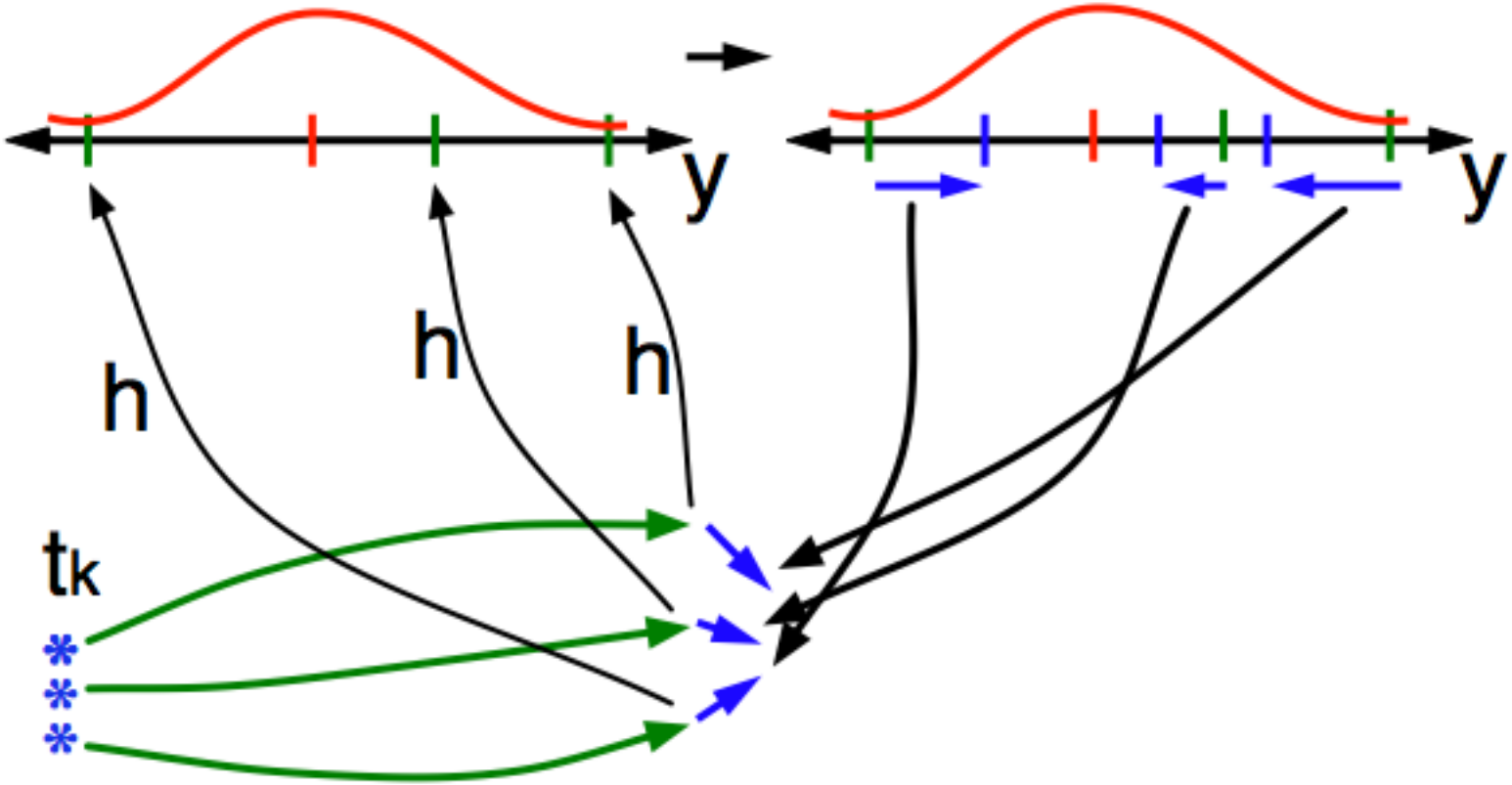
4. Find the increments for the prior observation ensemble.



Note: Difference between various ensemble filters is primarily in observation increment calculation.

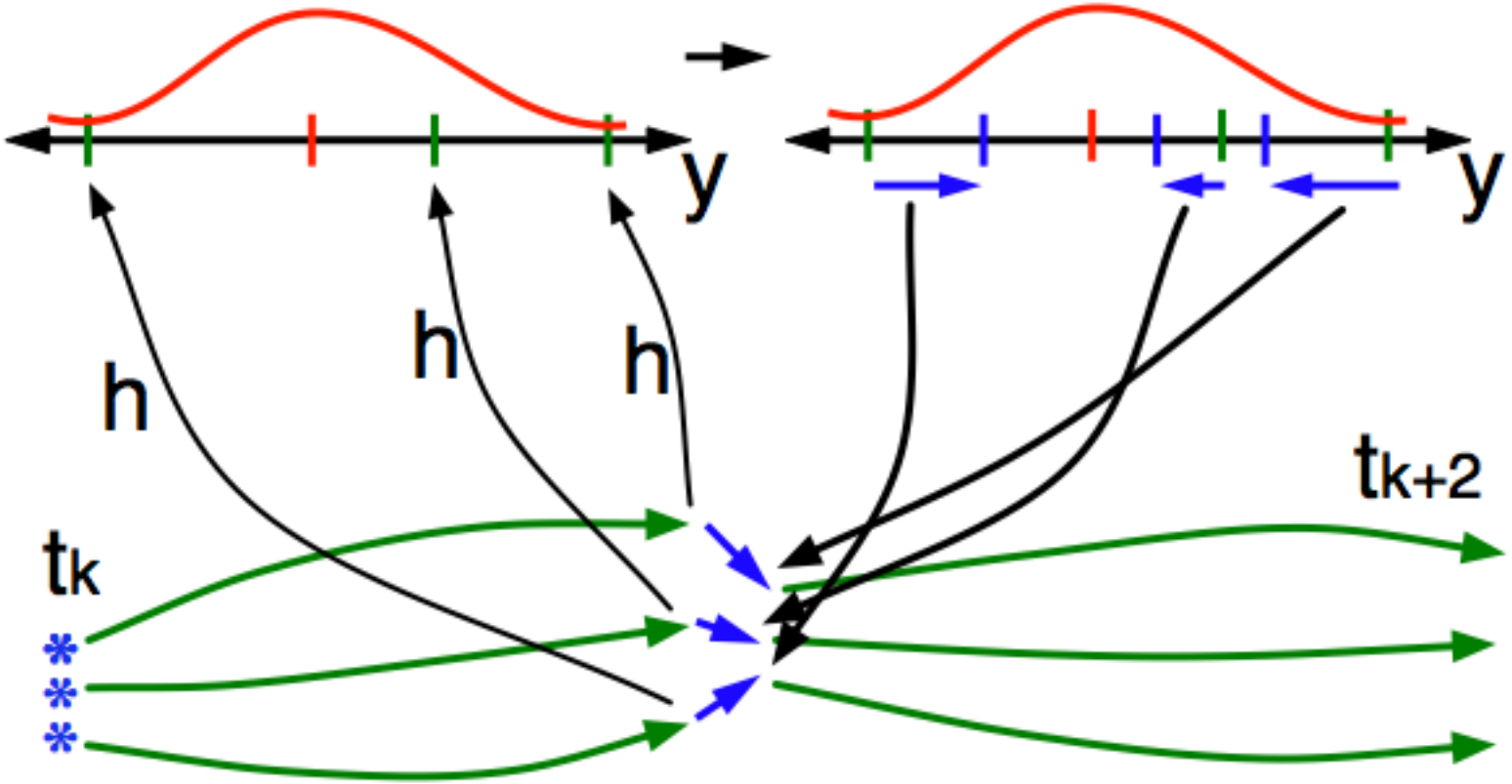
Ensemble filter for large geophysical models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



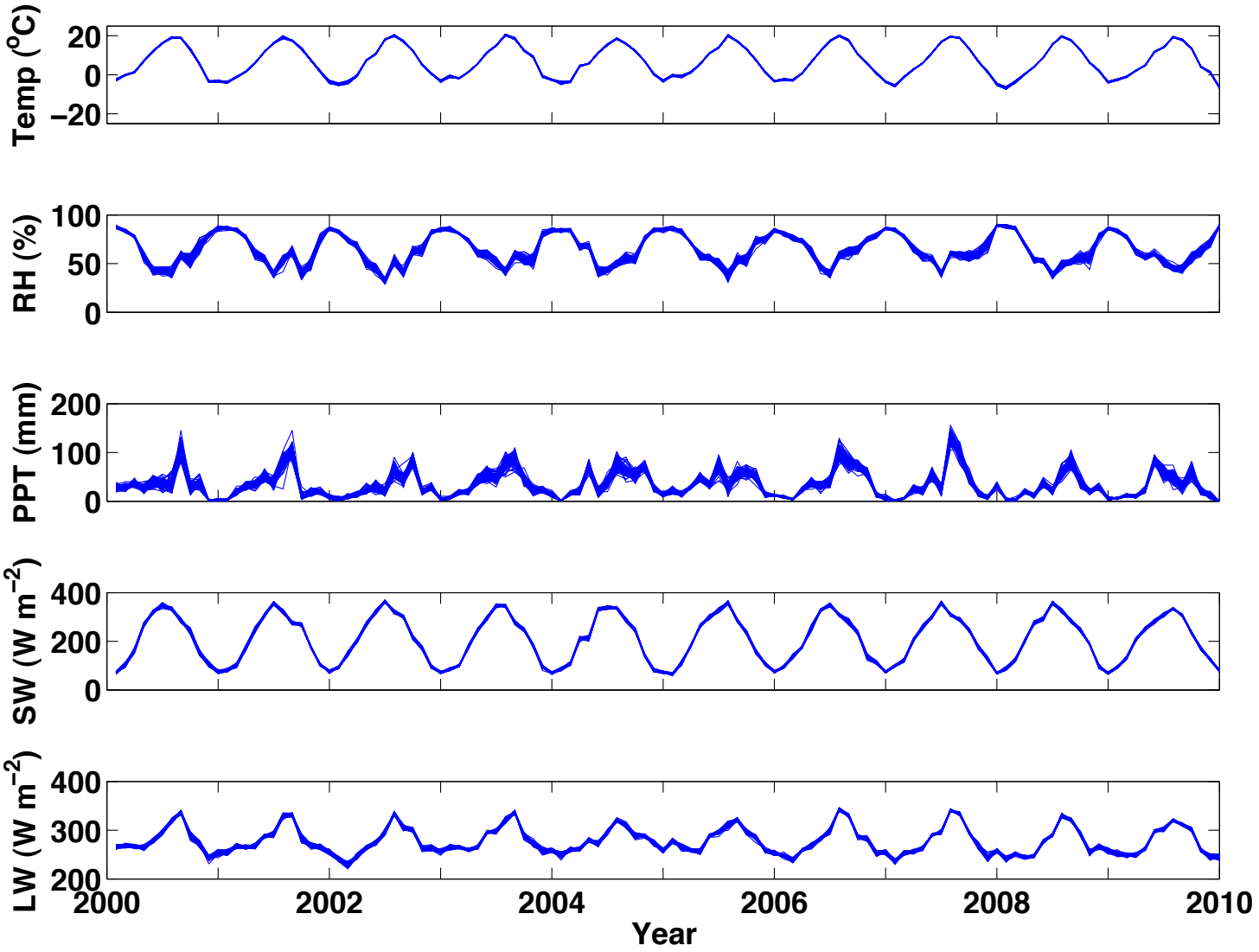
Ensemble filter for large geophysical models

6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …

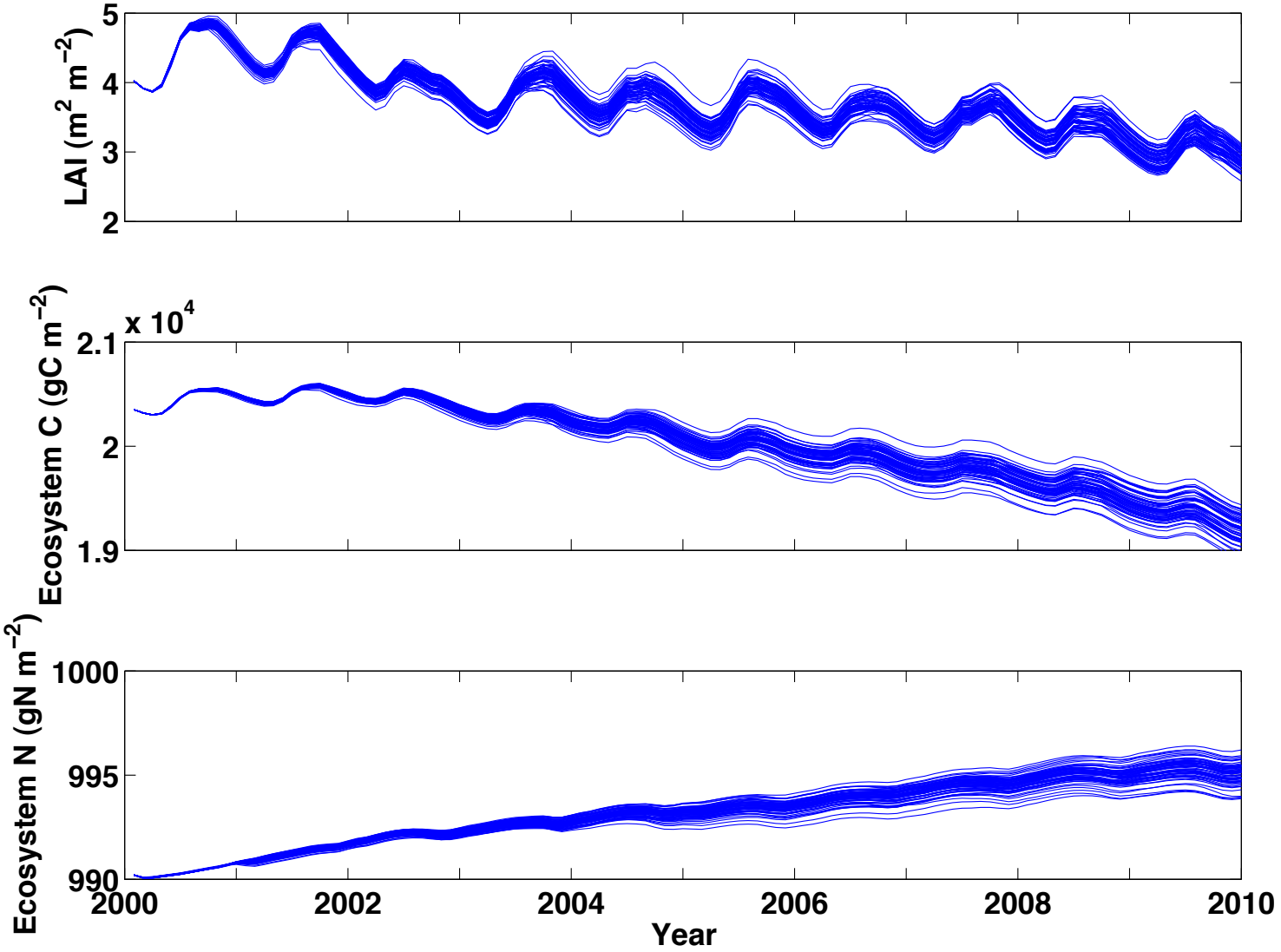


AN EXPERIMENT

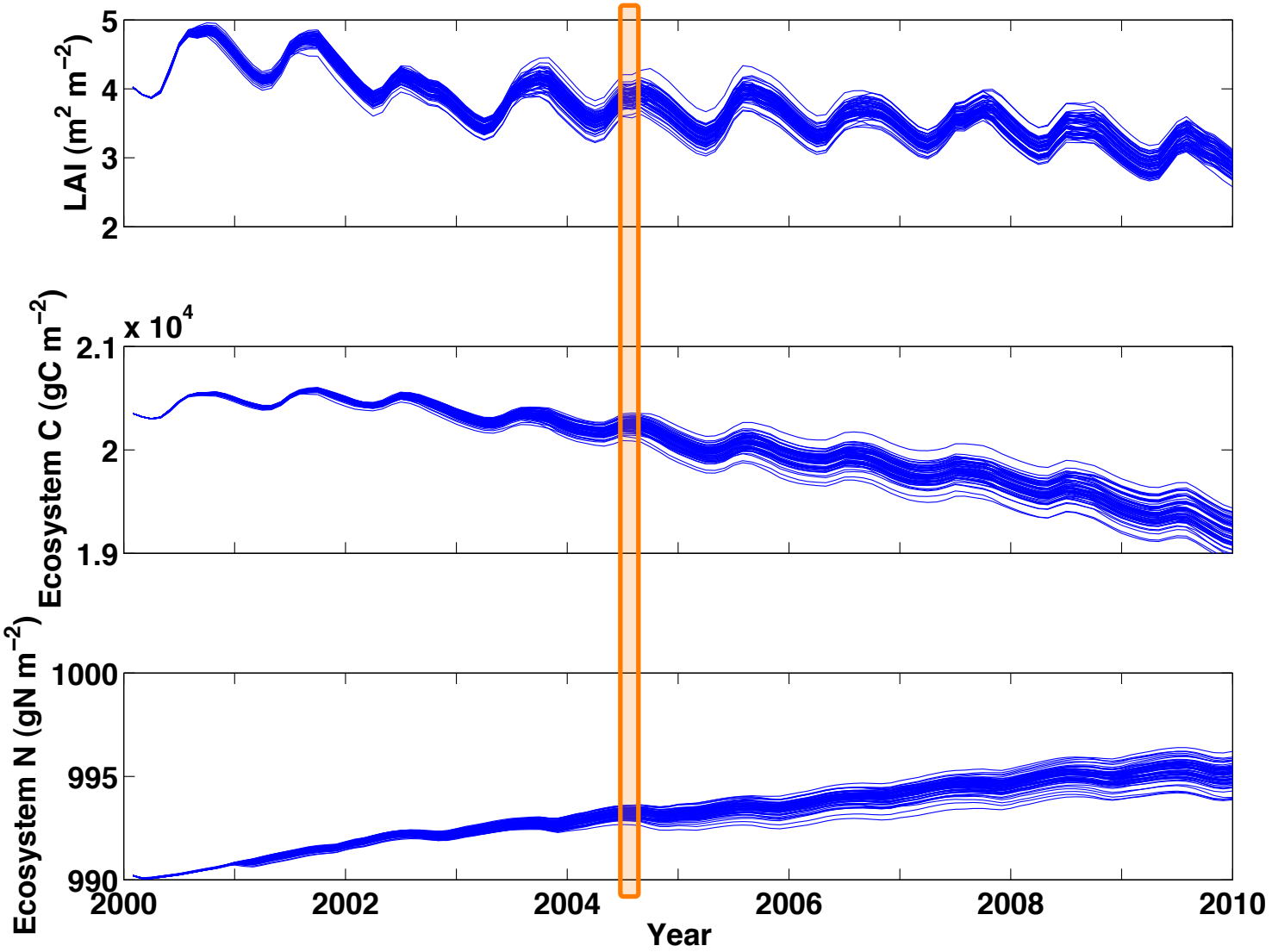
64 member climate ensemble at Niwot Ridge



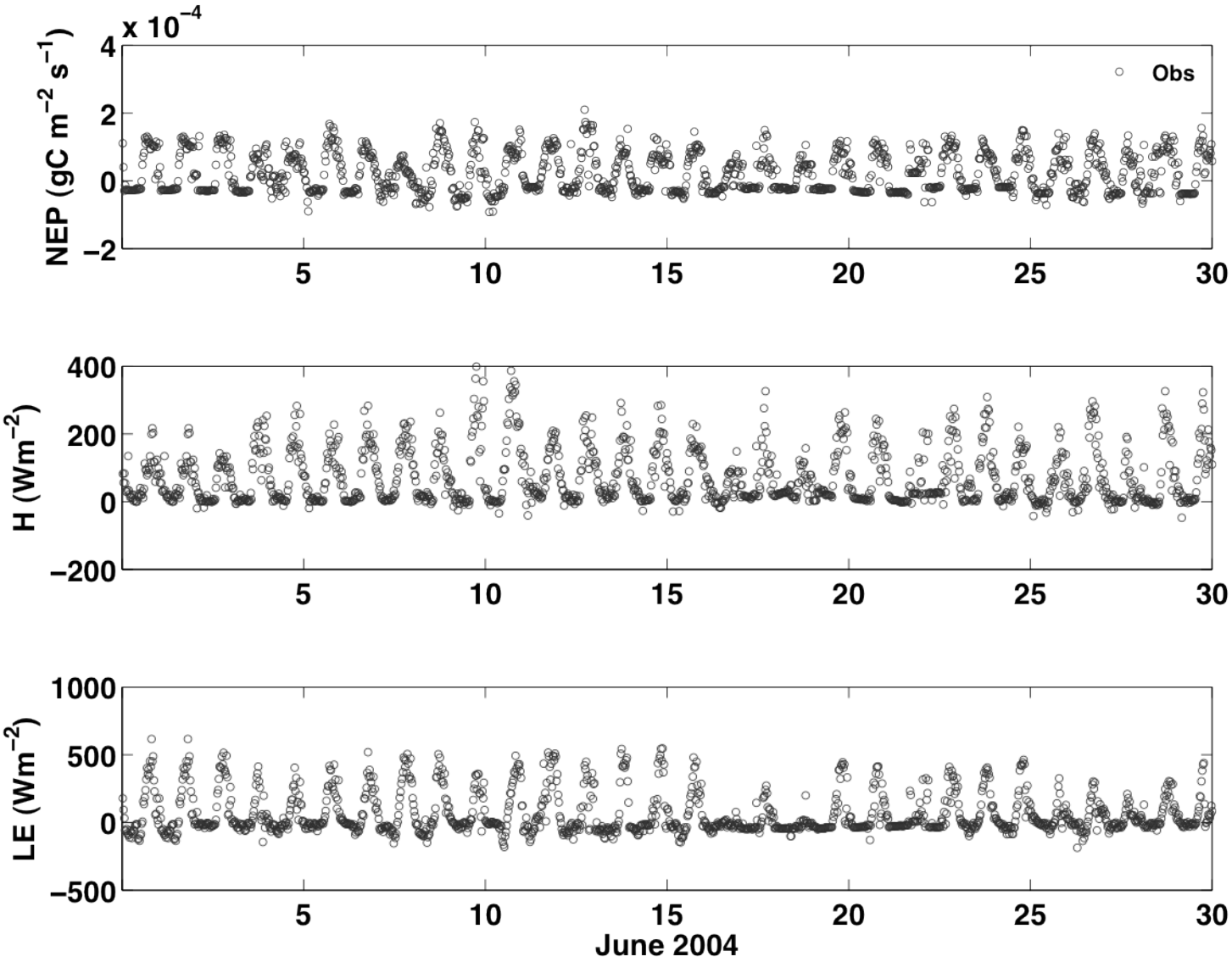
64 ensemble member free run of CLM



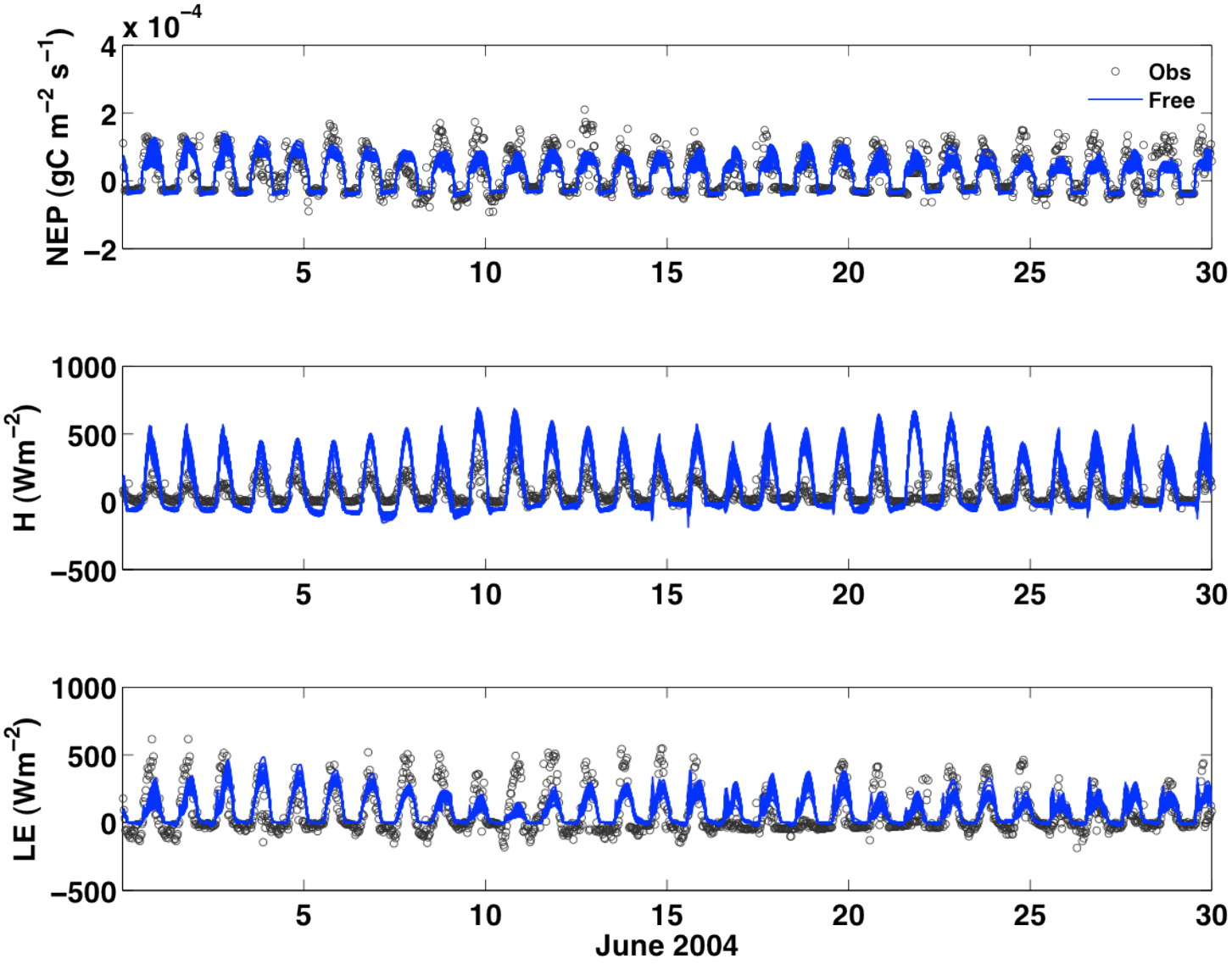
64 ensemble member free run of CLM



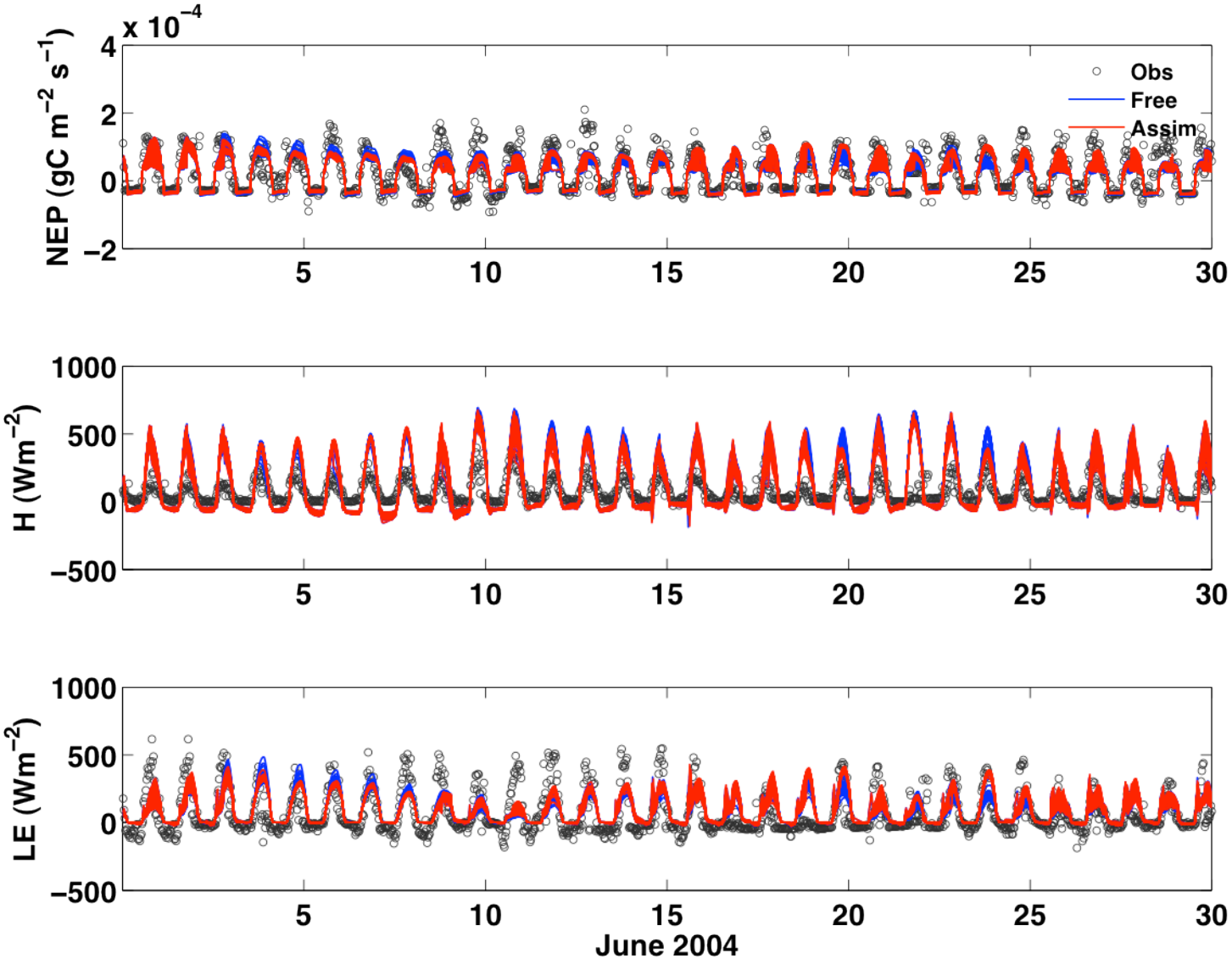
1 months of flux observations, June 2004



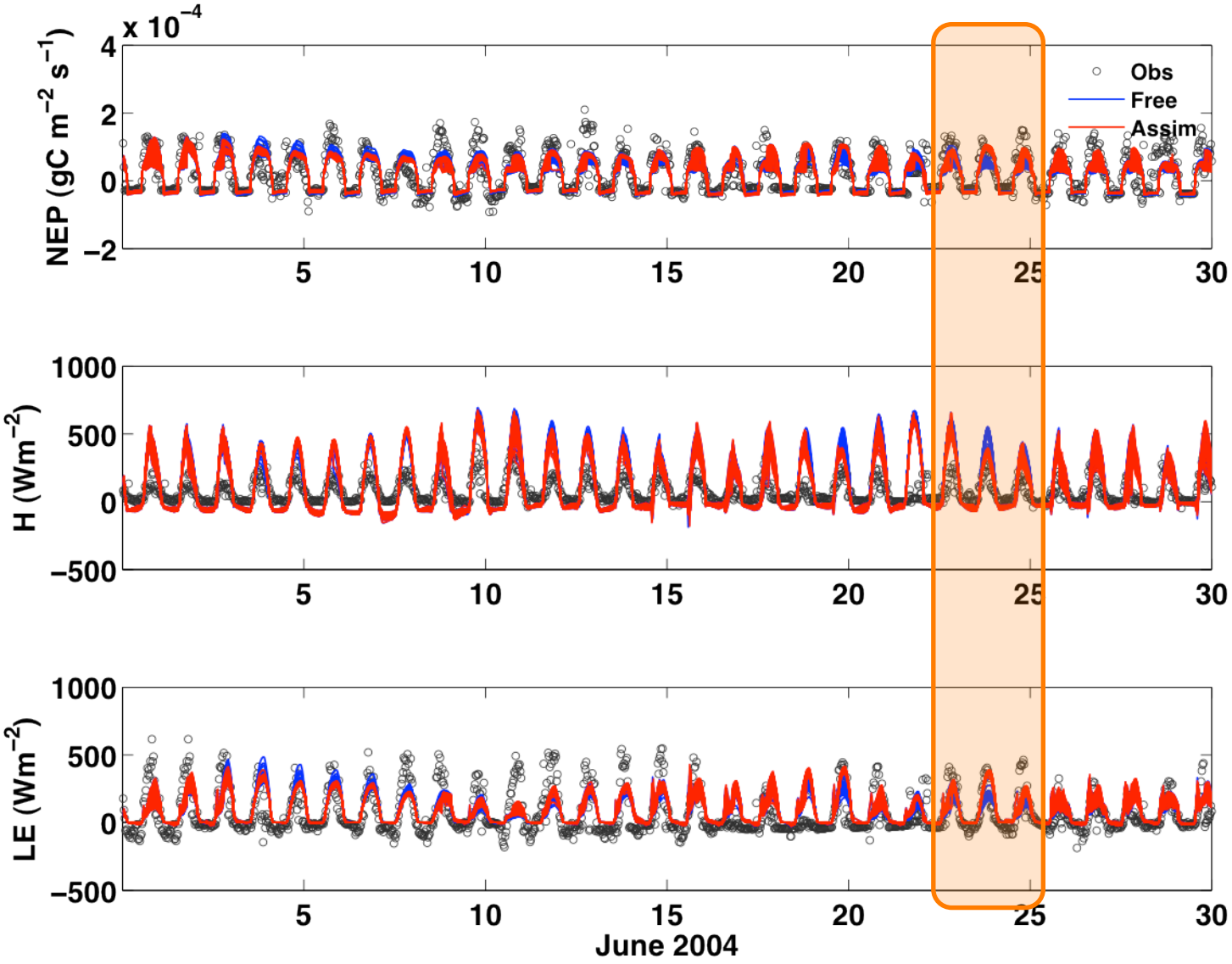
1 month free run



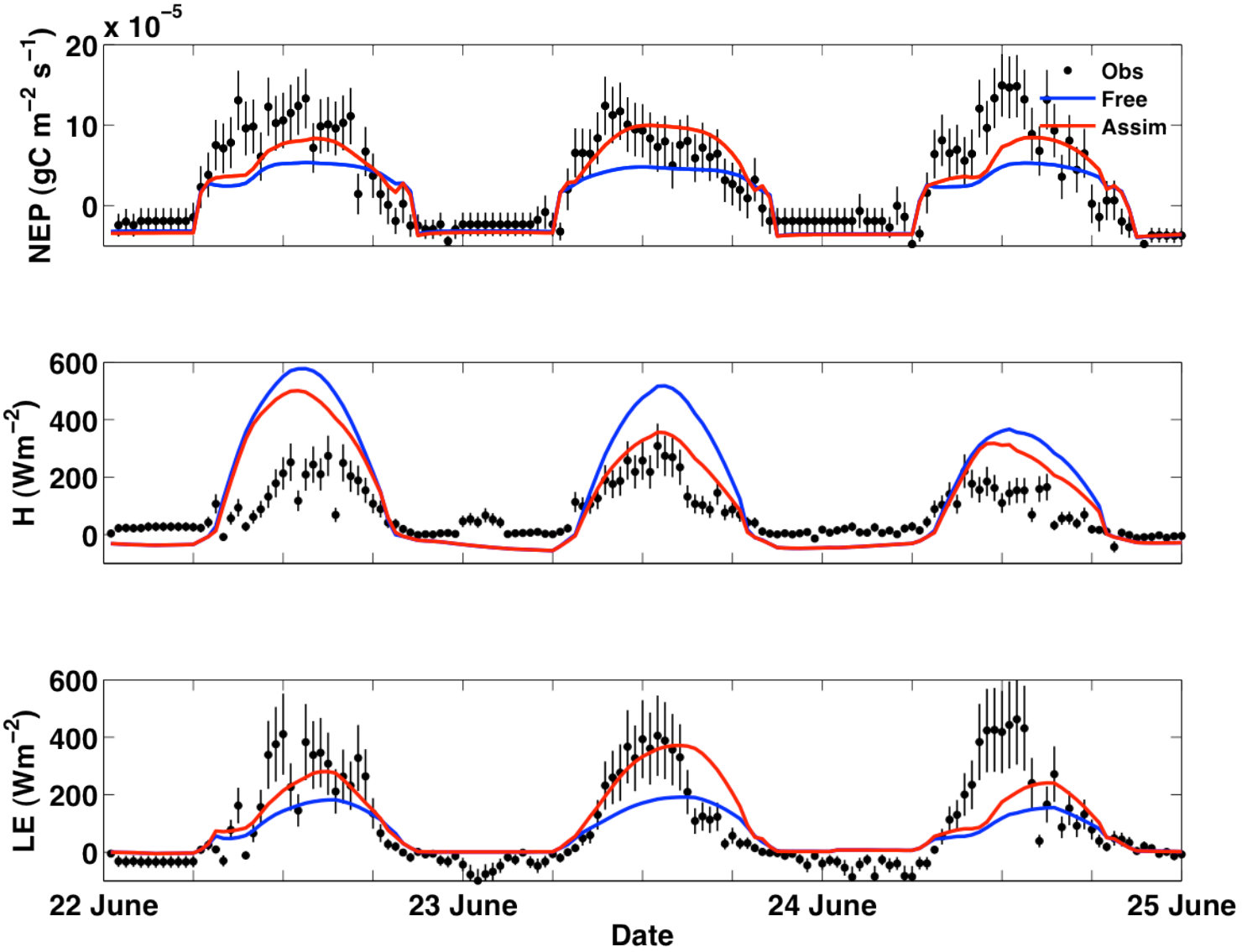
1 month assimilation run



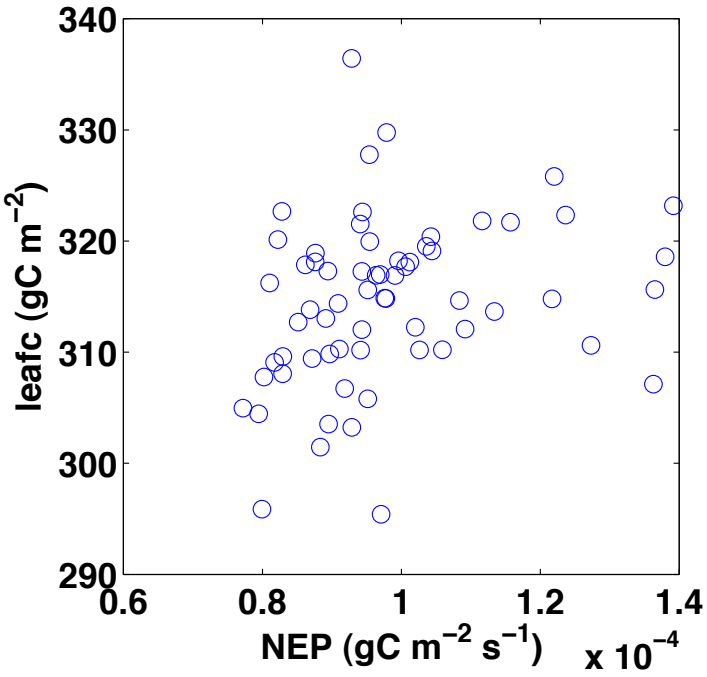
1 month assimilation run

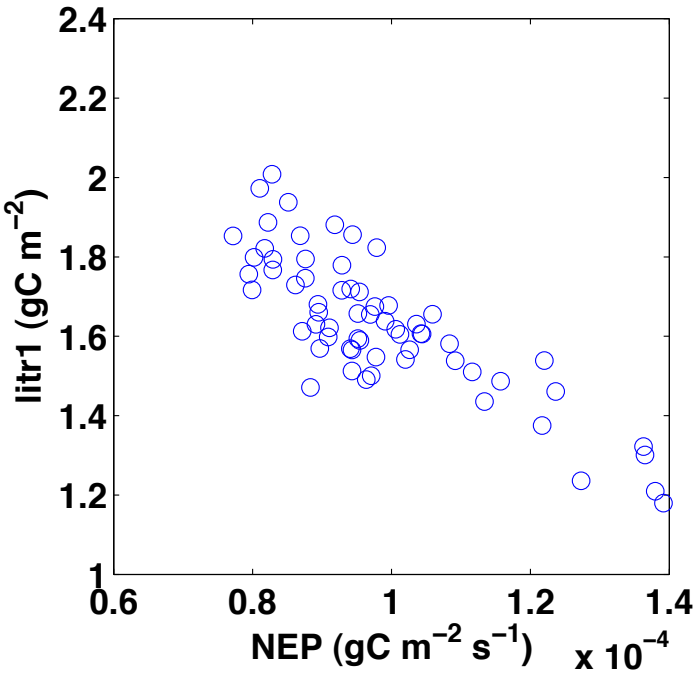


Looking at a few day – ensemble mean

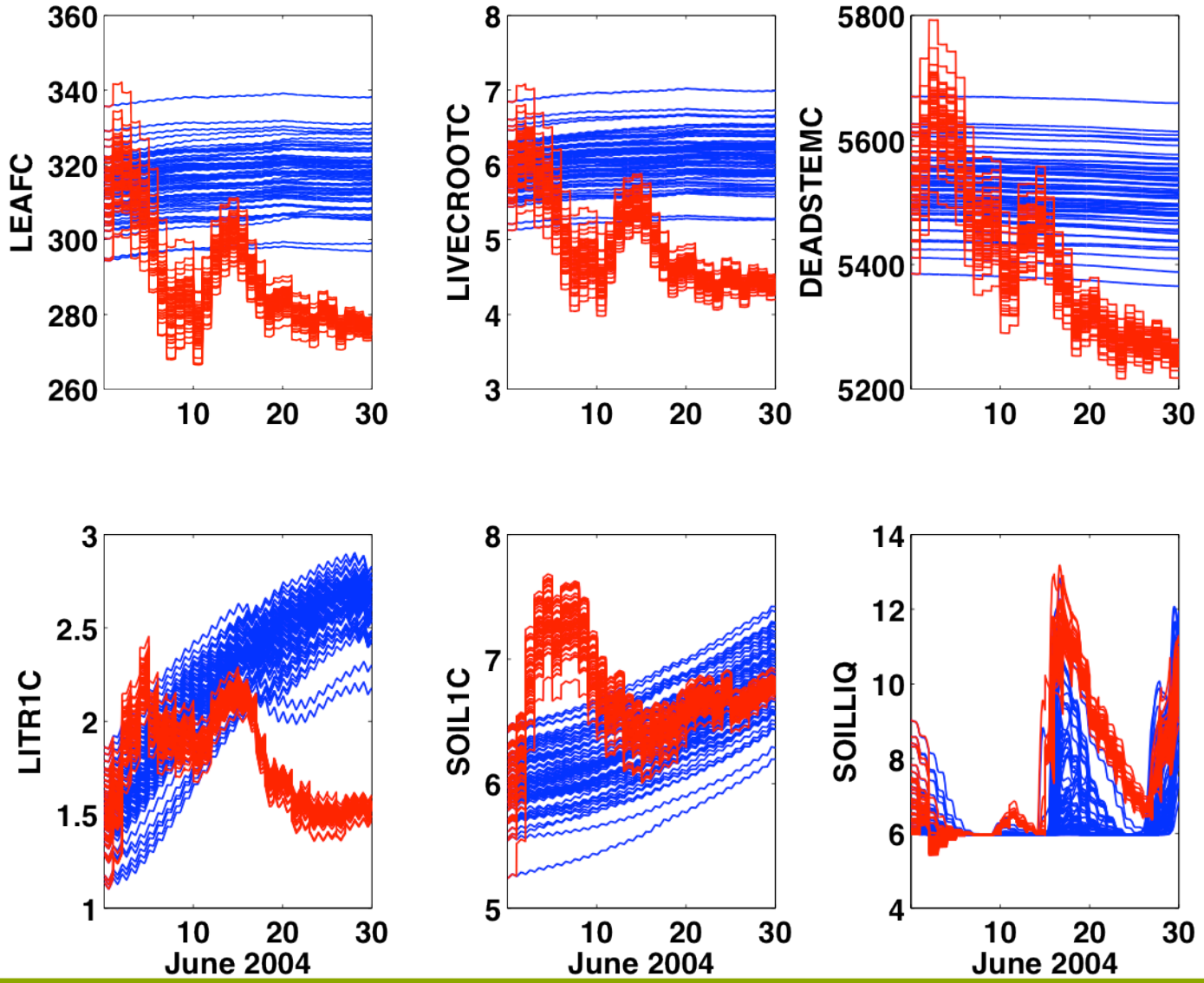


Unobserved v. Observed, 1300hrs, 3 June

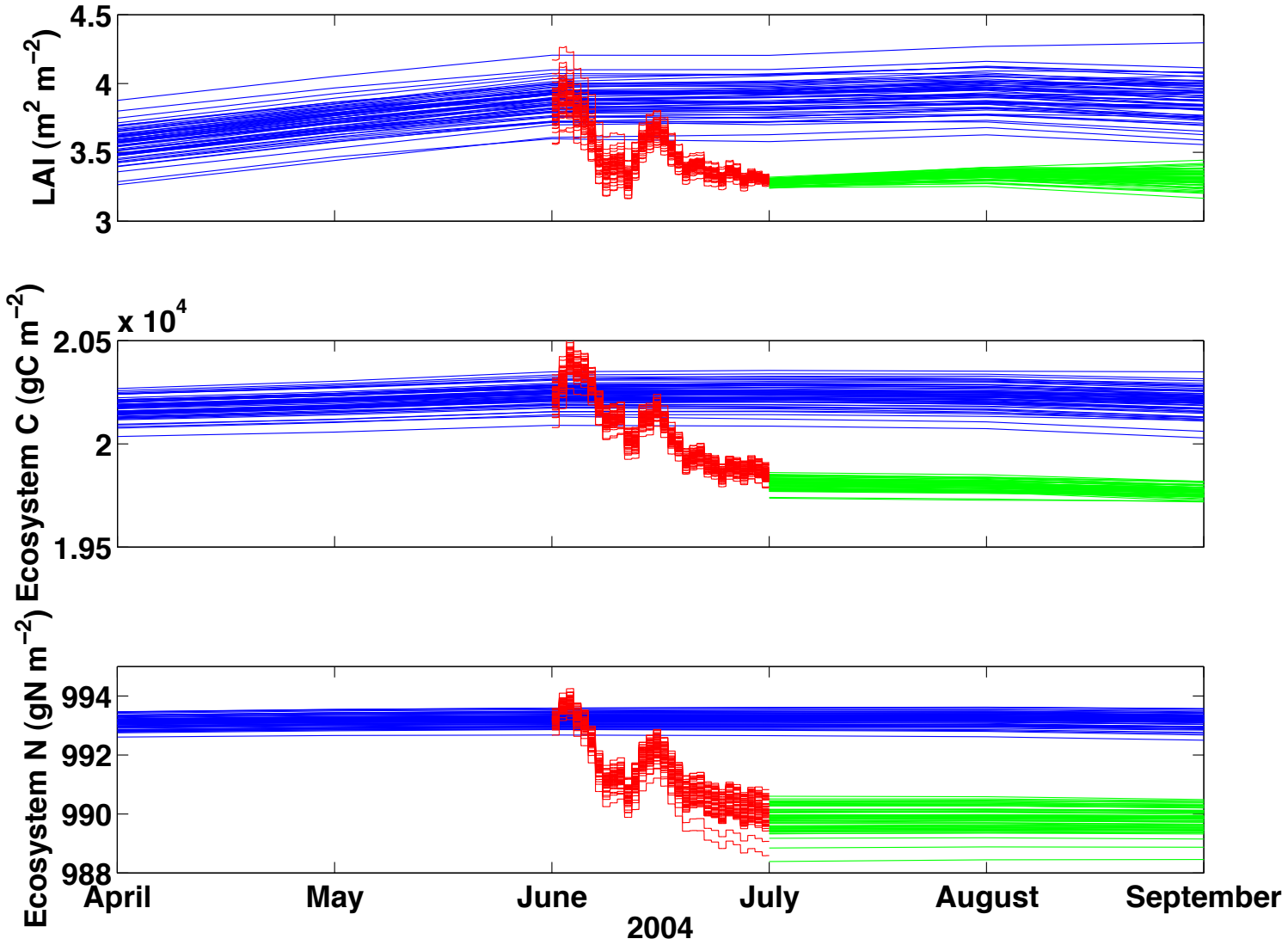




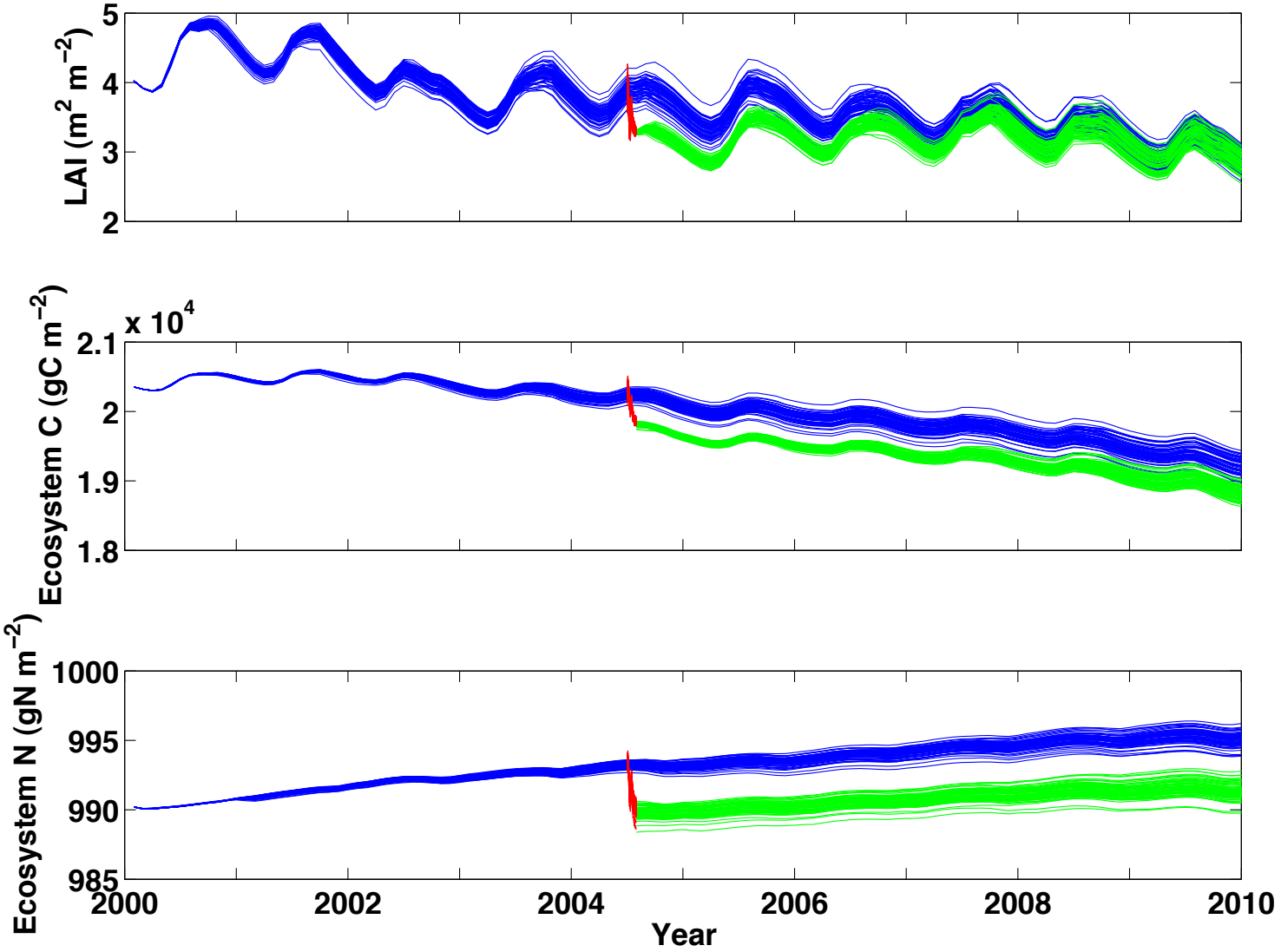
Unobserved variables



Effect on short-term forecast

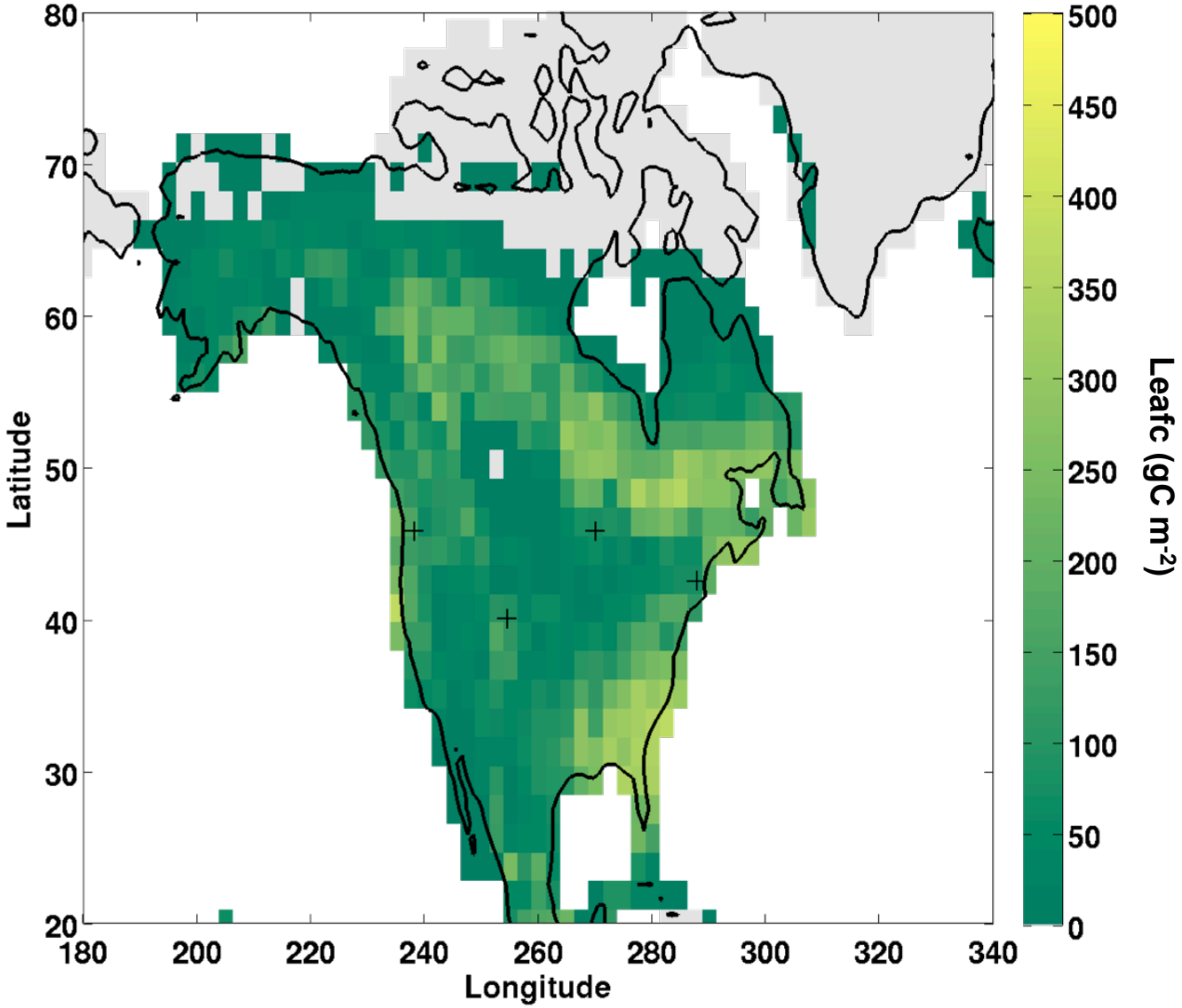


Effect on longer-term forecast

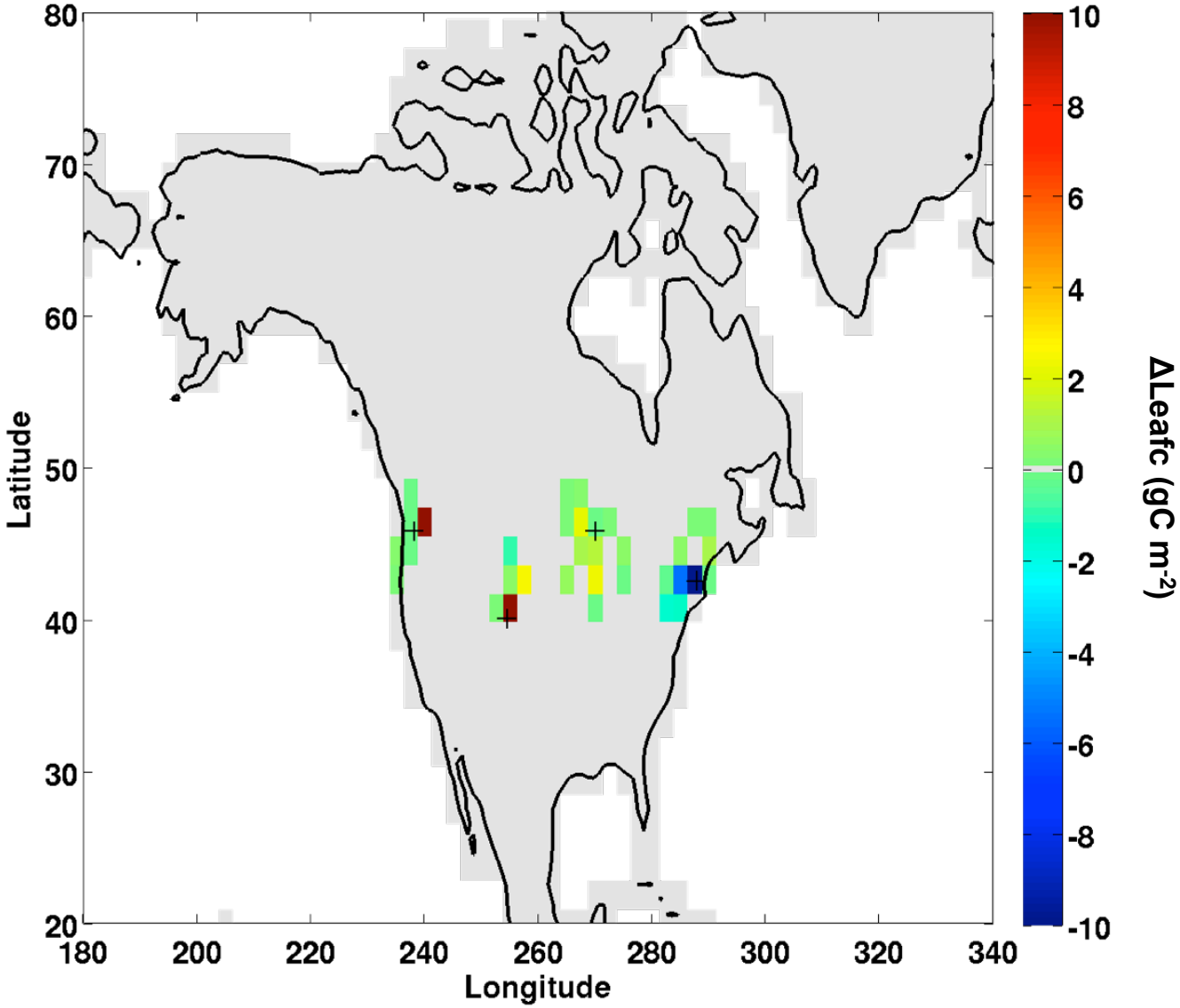


EXTRAPOLATION ACROSS SPACE

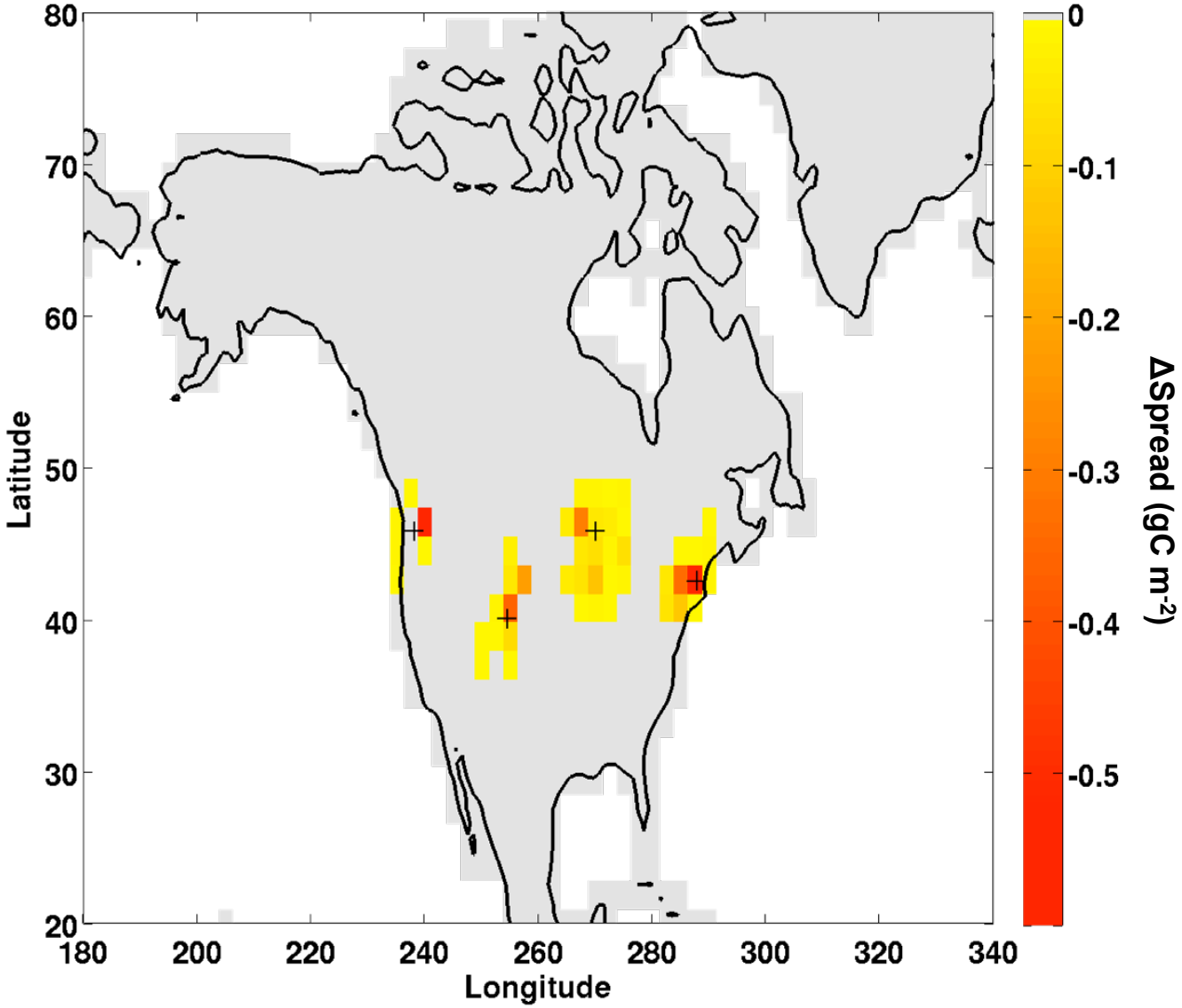
Continental scale *leafc*, 4 June 2000



Continental scale Δ *leafc*, 4 June 2000



Continental scale Δ spread, 4 June 2000



UNANSWERED QUESTIONS

Many big questions remain

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

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON Inc.

DART and CLM references

- 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. doi:10.1175/2009BAMS2618.1
- And many more here
  - http://www.image.ucar.edu/DAReS/Publications/index.php

- 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.
- Lawrence, D.M., et al., 2011: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Sys.*, **3**, DOI: 10.1029/2011MS000045.
- And many more here
  - http://www.cesm.ucar.edu/models/cesm1.0/clm/clm_bibliography.htm

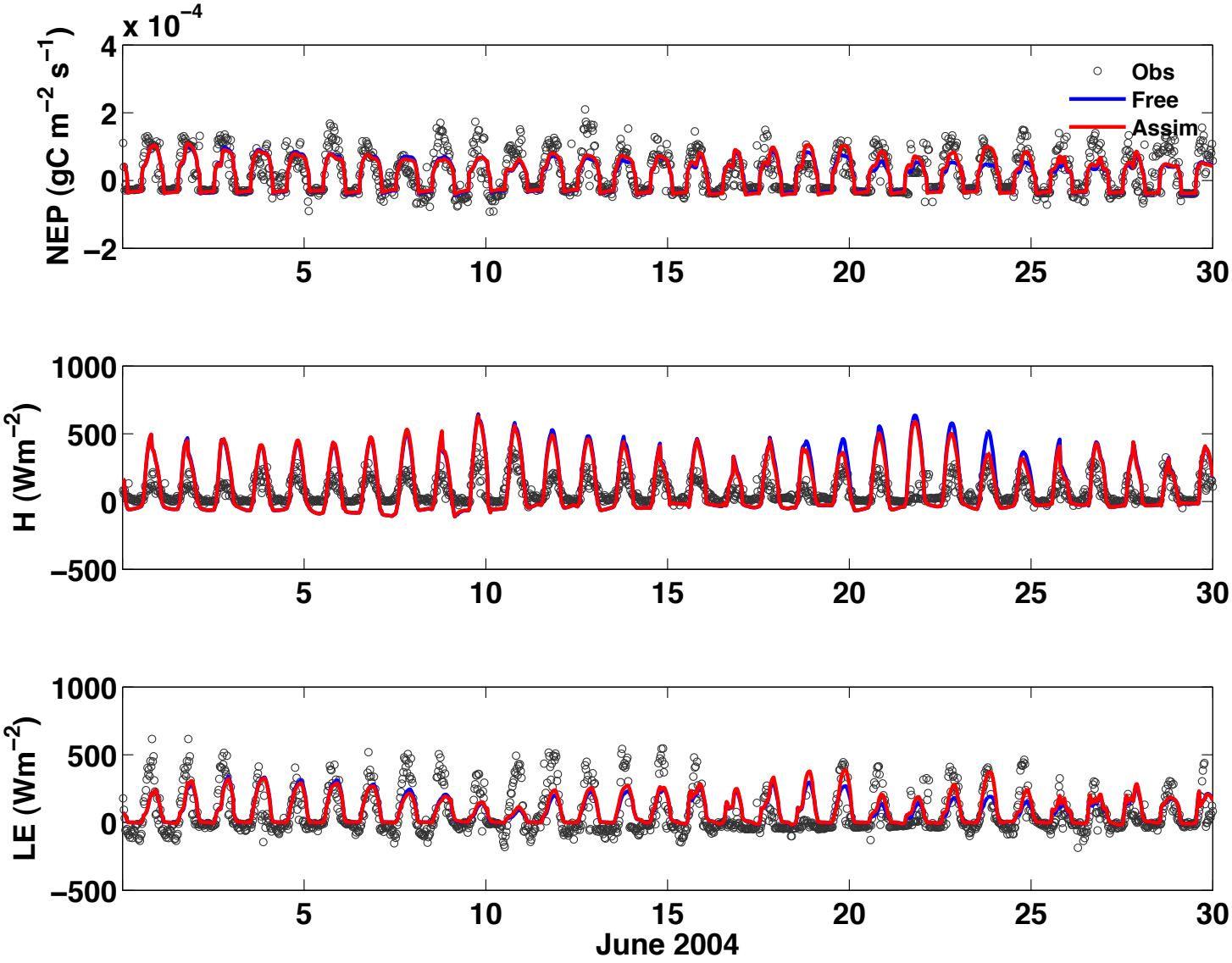
Some philosophical points…

- Estimating parameter values is common approach in these ecosystem models
- Is estimating model states a useful thing to be doing?
- Admitting the models are “wrong”
- But you’re not trying to improve them…
- Depends on your questions…

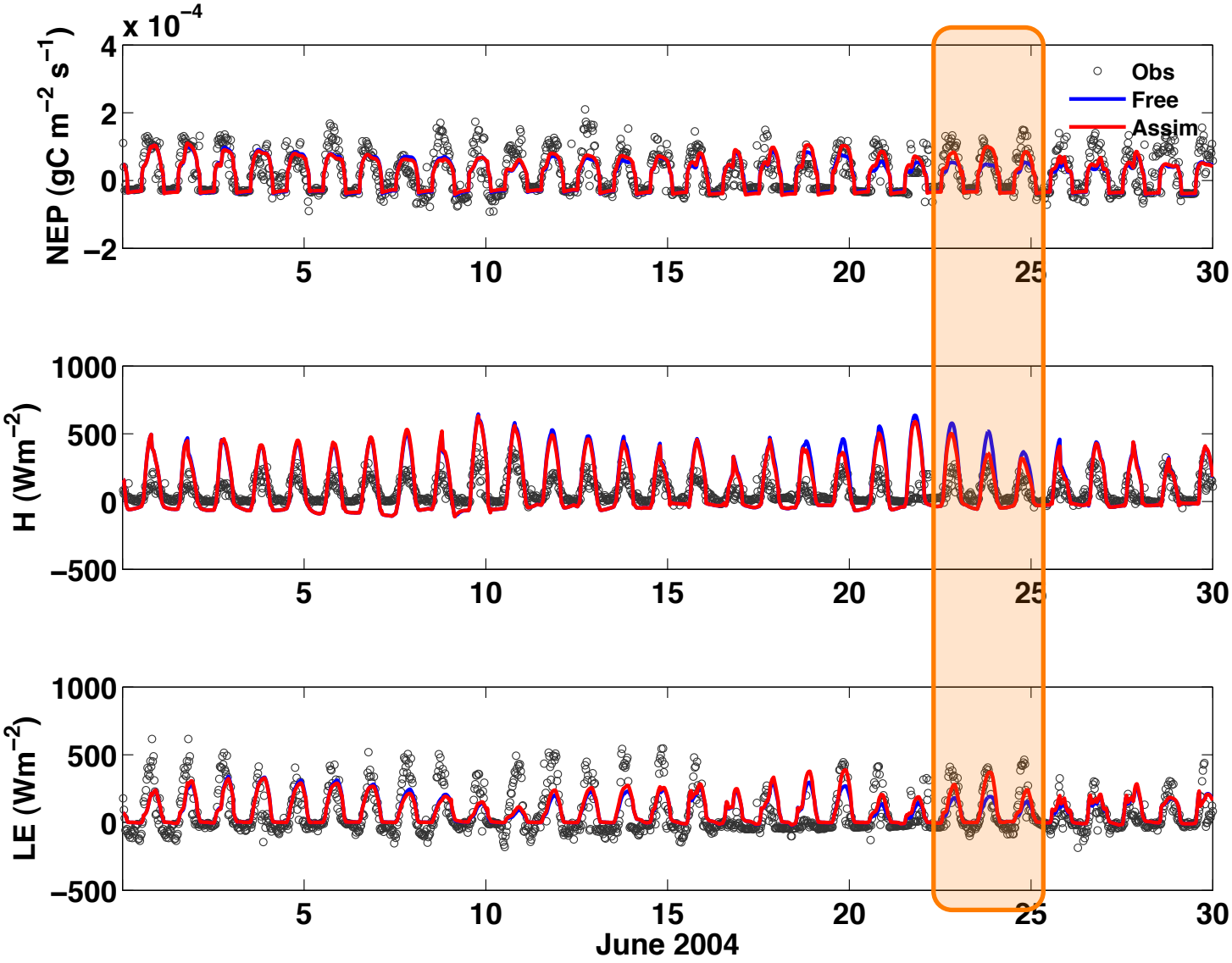
Is DA different for NWP and ecosystem models?

	Data Assimilation in NWP	Data Assimilation in CLM
Main objective	Forecast improvement	Process understanding Regional quantification Forecasting
Dynamics	Physics – essentially well known from first principles	Physical, biological, chemical – Only partially known, empirical relationships
Observations	High spatial and temporal density	Very different spatial and temporal characteristics
Mathematical problem	Optimization of initial conditions	Initial value problem (e.g. pools) Boundary conditions (e.g. fluxes) Parameter optimization

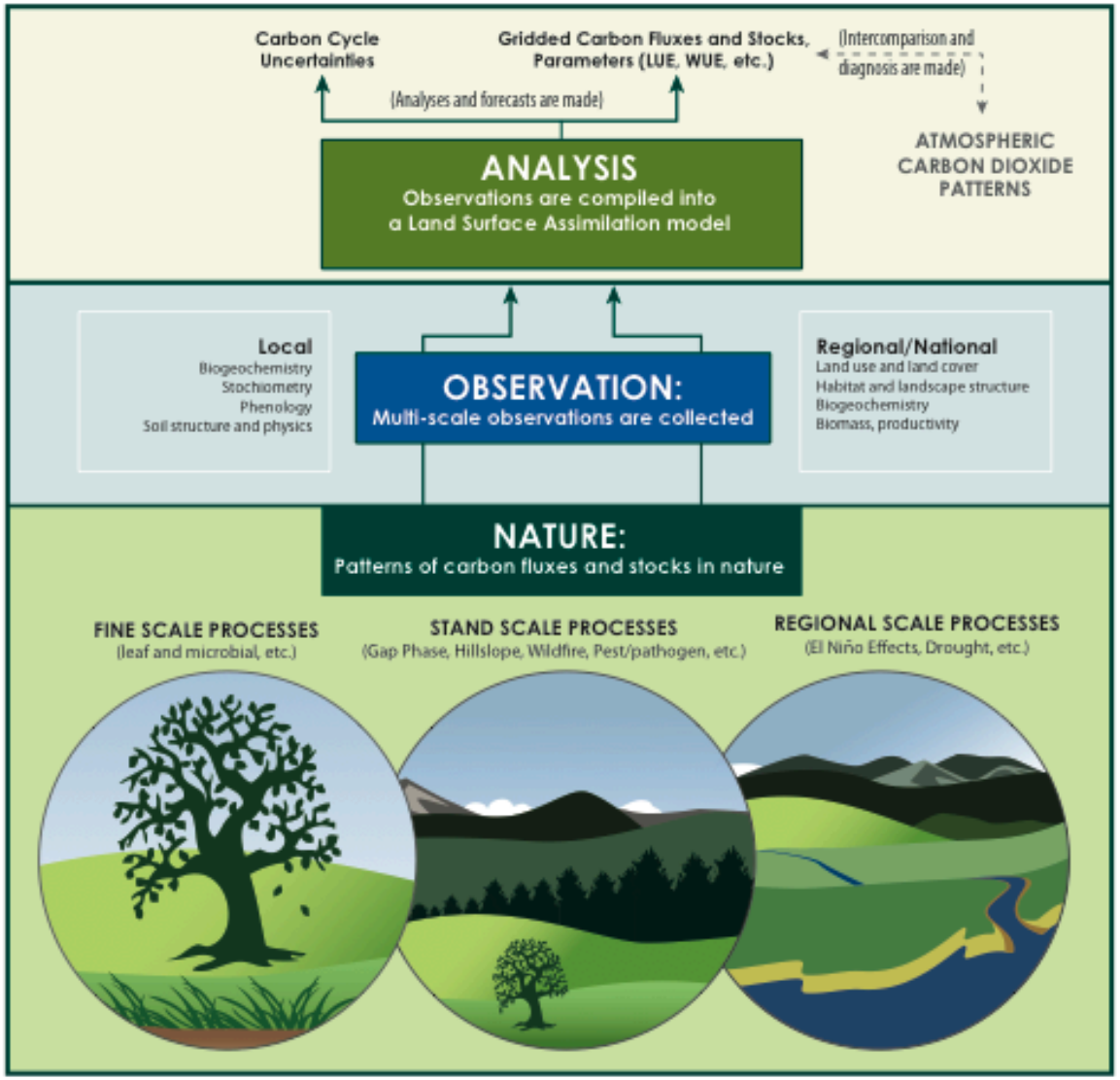
1 month assimilation run – ensemble mean



1 month assimilation run – ensemble mean



Extrapolating observations to continental scale



Data Assimilation is able to…

- Provide estimates of
  - state variables
  - initial conditions
  - Parameters
- And quantifies their uncertainties
- Helping to select between alternative model structures
- Providing a quantitative basis to evaluate sampling strategies for future experiments and observations
- This will enable improvements to BOTH models and forecasts