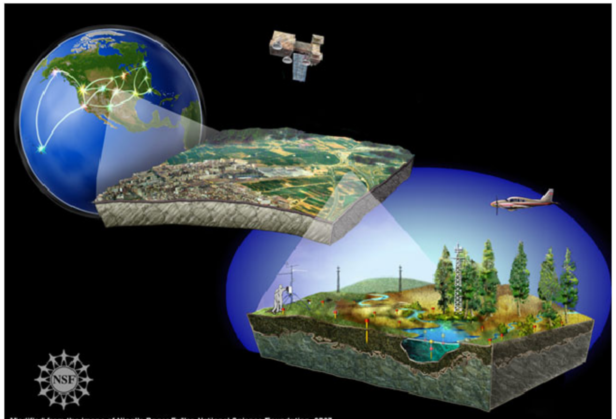
USING ENSEMBLE DA TO INFORM CARBON DYNAMICS IN THE COMMUNITY LAND MODEL

Andy Fox[1], Tim Hoar[2]

1. National Ecological Observatory Network
2. National Center for Atmospheric Research

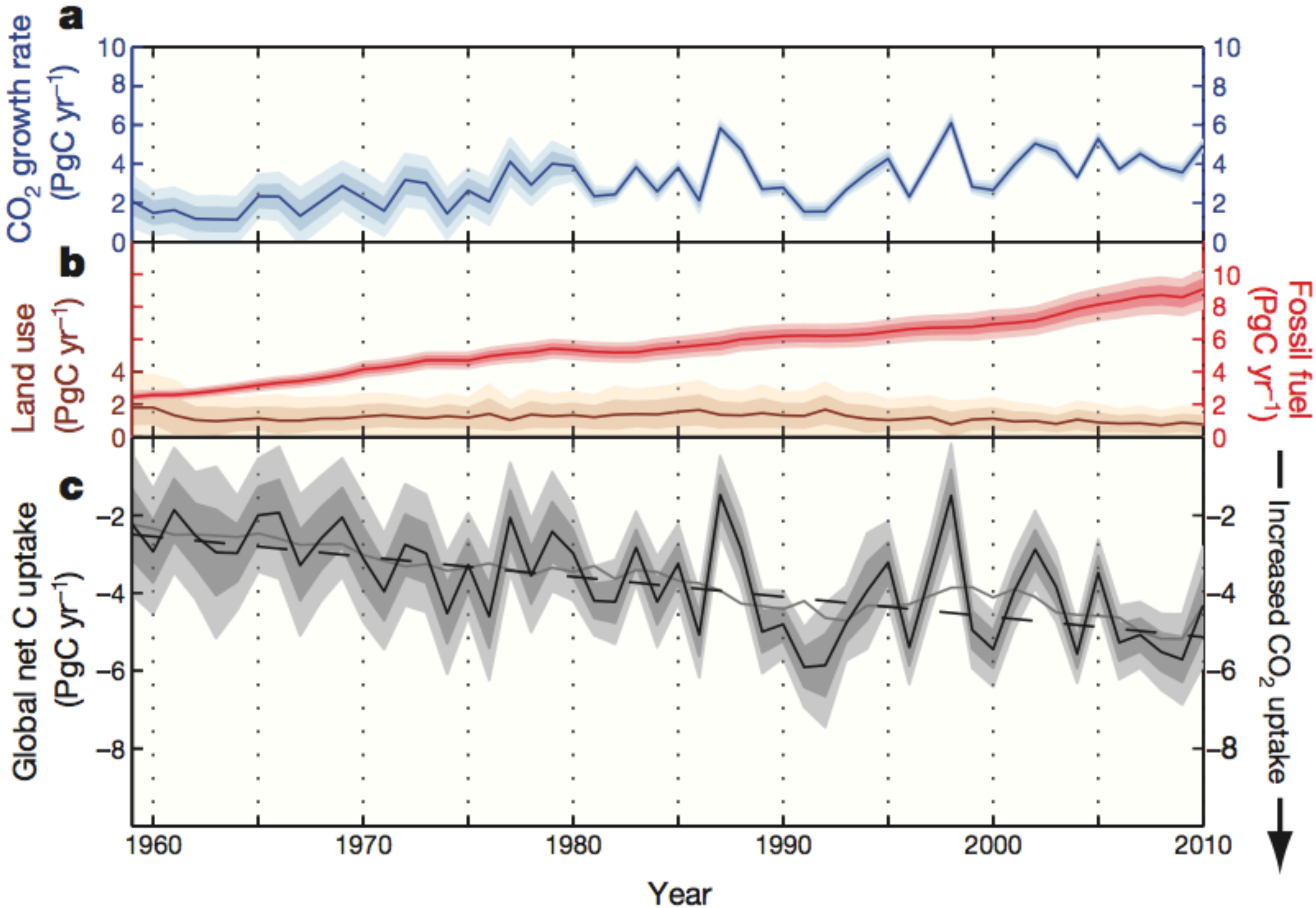


Modified from the image of Nicolle Rager Fuller, National Science Foundation, 2007

Outline

1. Introduction
2. Data
3. Community Land Model
4. Data Assimilation Research Testbed
5. Perfect Model Experiments
6. Some Real Data
7. Unanswered Questions

Coupled Climate-Carbon Cycle models



Ballantyne et al., 2012

Uncertainty in Coupled Climate-CC Models

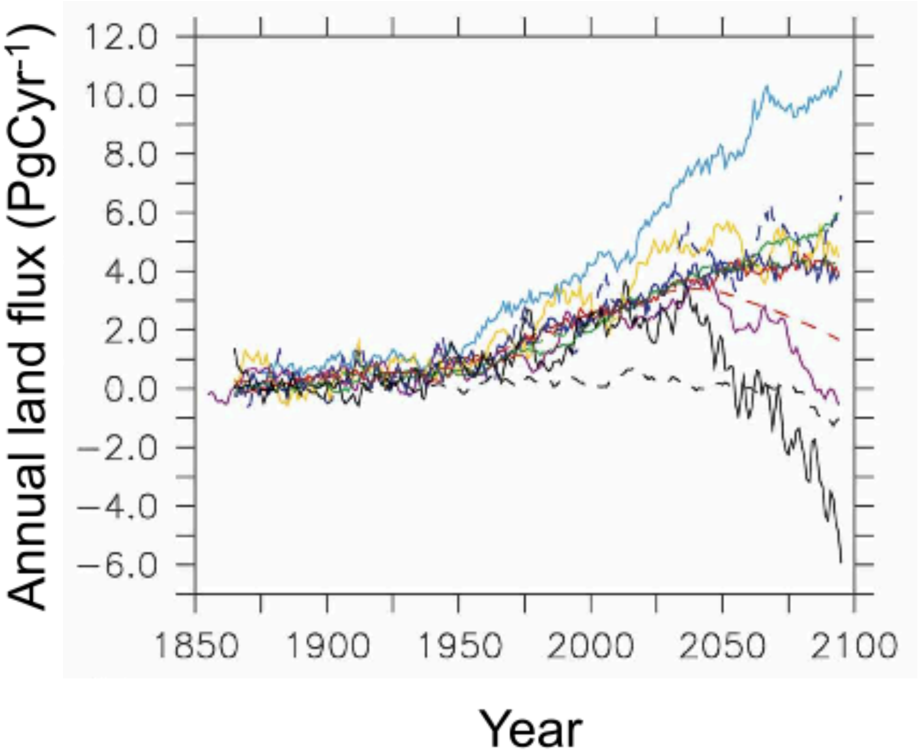
- This uncertainty stems from
  - i. **Structural uncertainty**
  - ii. **Parameter uncertainty**
  - iii. **Initial conditions uncertainty**
  - iv. **Boundary conditions uncertainty**

VOLUME 19 JOURNAL OF CLIMATE 15 JULY 2006

2006

Climate–Carbon Cycle Feedback Analysis: Results from the C^{4}MIP Model Intercomparison

P. FRIEDLINGSTEIN,[a] P. COX,[b] R. BETTS,[c] L. BOPP,[a] W. VON BLOH,[d] V. BROVKIN,[d] P. CADULE,[e] S. DONEY,[f] M. EBY,[g] I. FUNG,[h] G. BALA,[i] J. JOHN,[h] C. JONES,[c] F. JOOS,[j] T. KATO,[k] M. KAWAMIYA,[k] W. KNORR,[l] K. LINDSAY,[m] H. D. MATTHEWS,[g,n] T. RADDATZ,[o] P. RAYNER,[a] C. REICK,[o] E. ROECKNER,[p] K.-G. SCHNITZLER,[p] R. SCHNUR,[p] K. STRASSMANN,[j] A. J. WEAVER,[g] C. YOSHIKAWA,[k] AND N. ZENG[q]



Uncertainty in Coupled Climate-CC Models

- This uncertainty stems from
  i. **Structural uncertainty**
  ii. **Parameter uncertainty**
  iii. **Initial conditions uncertainty**
  iv. **Boundary conditions uncertainty**

- Need to find (new) ways to use (new) observations to:
  – Evaluate
  – Benchmark
  – Constrain
  – Assimilate

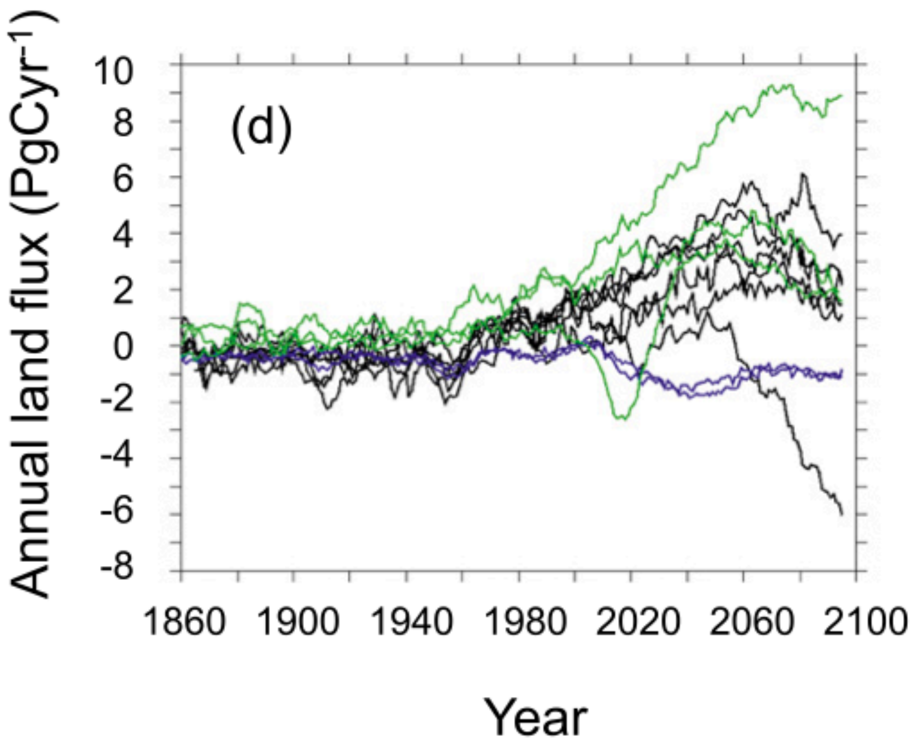
15 January 2014 FRIEDLINGSTEIN ET AL. 511

SPECIAL C4MIP COLLECTION

2014

Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks

Pierre Friedlingstein,* Malte Meinshausen,+ Vivek K. Arora,# Chris D. Jones,@ Alessandro Anav,* Spencer K. Liddicoat,@ and Reto Knutti&

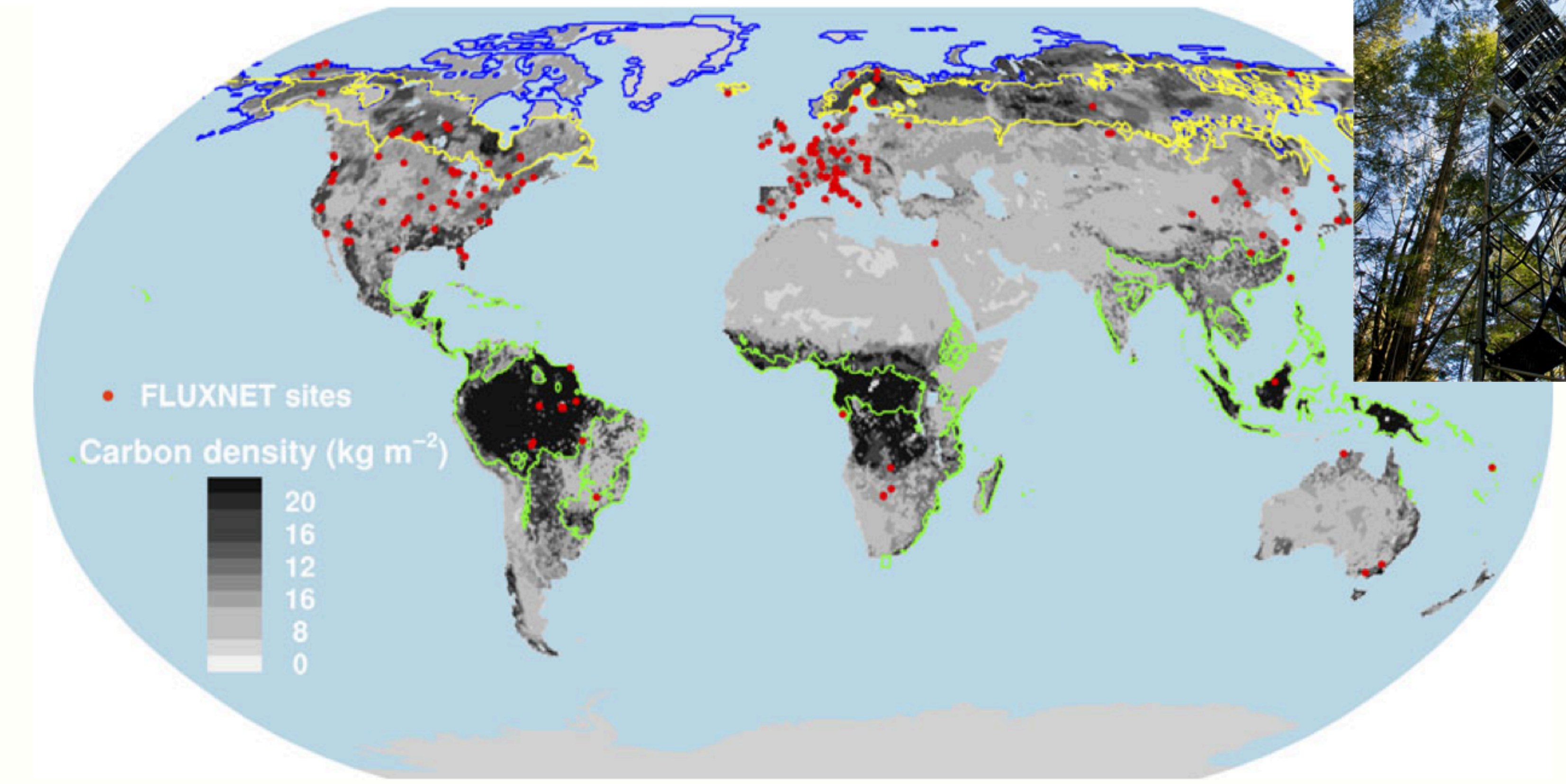


Is DA different for NWP and CC models?

	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

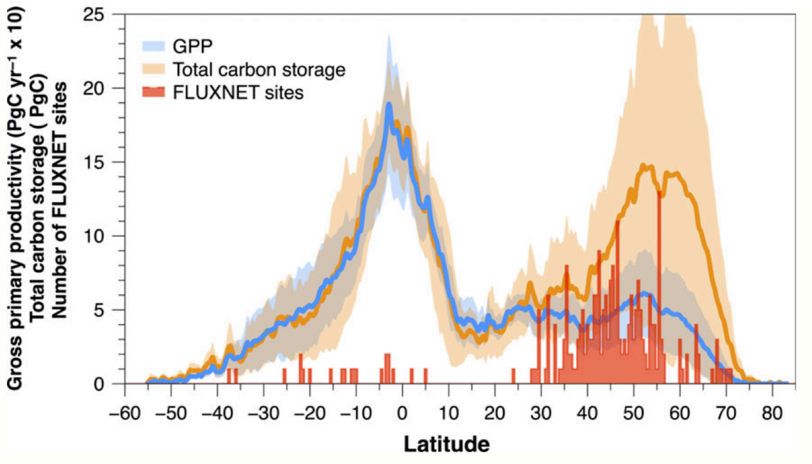
DATA

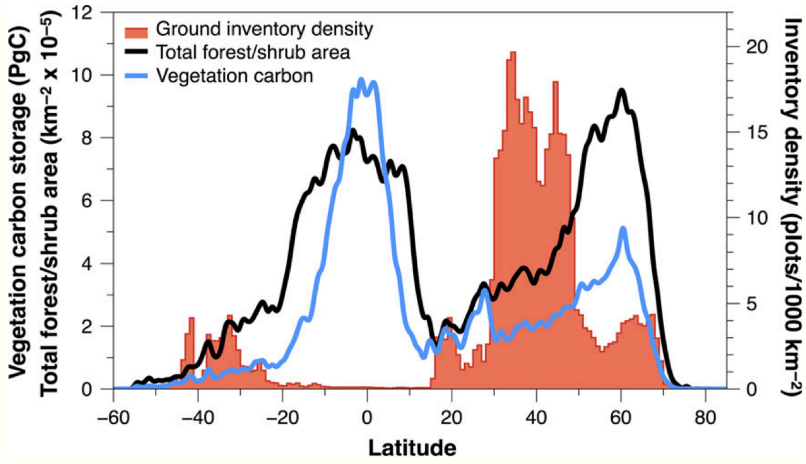
In-situ Flux tower observations

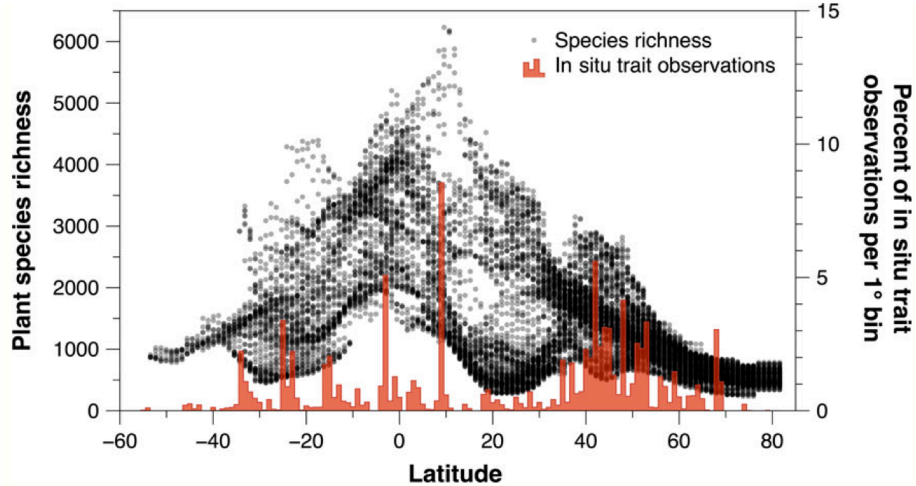


Schimel *et al.*, 2015

In-situ vegetation observations



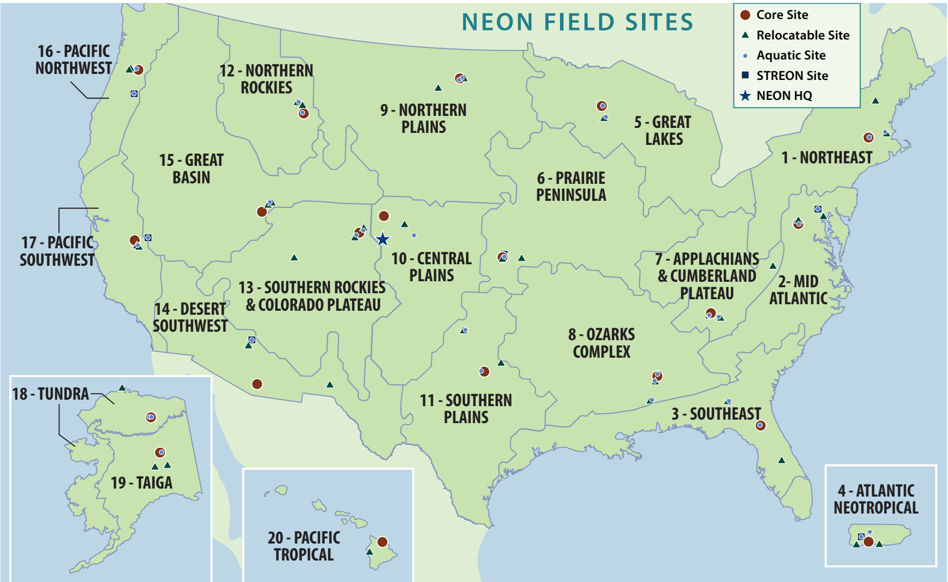




Schimel *et al.*, 2015

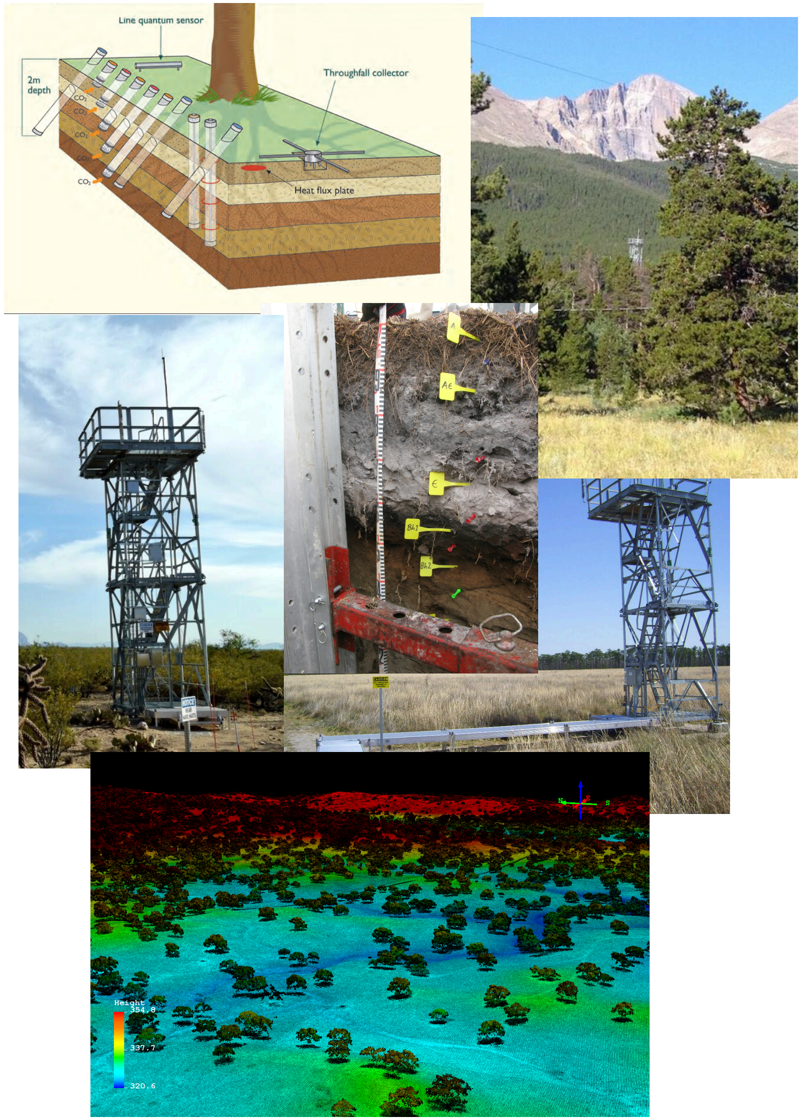
National Ecological Observatory Network

- Collect and openly distribute data on the drivers of and responses to ecological change
- Continental scope and 30-year time horizon
- Standardized methods of data collection, high investment in QA/QC, and calibration



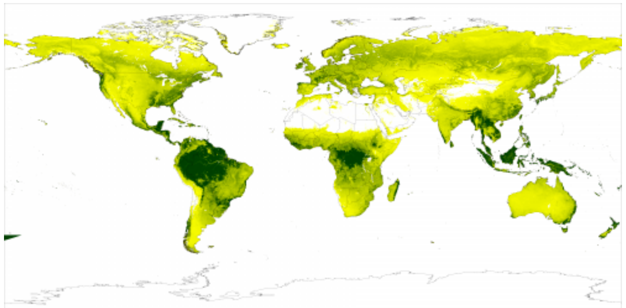
Carbon cycle Observations

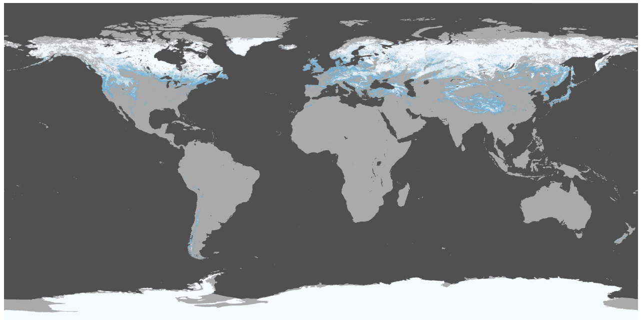
- Many relevant observations
- Some standard, some less common
  - Eddy covariance fluxes of energy, water and carbon
  - Profiles of soil temperature and moisture, and soil respiration
  - Profiles of soil carbon and nitrogen pools
  - NPP, litterfall and fine root turnover from minirhizotrons
  - Profiles of CO_2 and H_2O vapor isotopes
  - Soil microbial biomass, diversity & functional composition
  - Lidar and hyperspectral derived biomass, leaf area and canopy chemistry at <1m resolution over 100s km^2



Remote Sensing (Products)

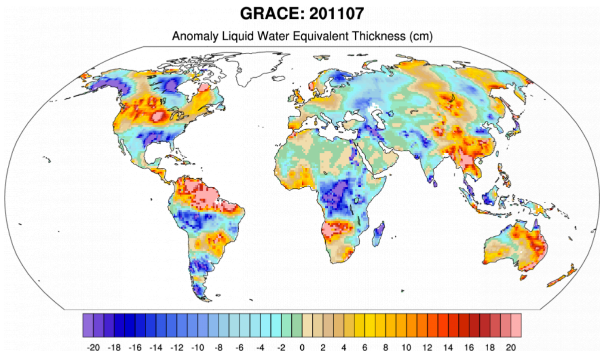
MODIS LAI

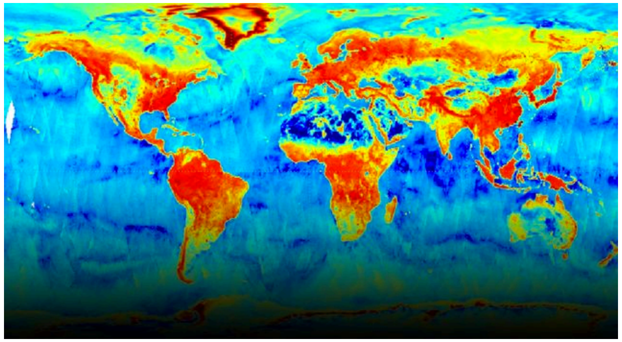




MODIS SCF

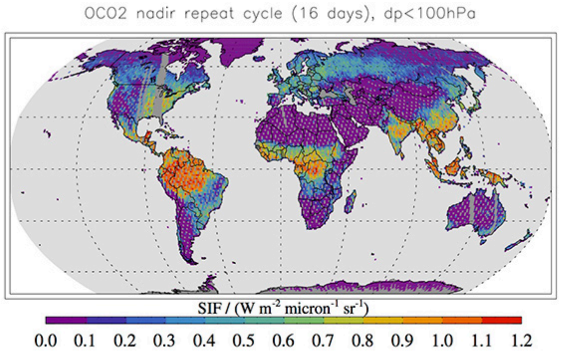
GRACE anomalies





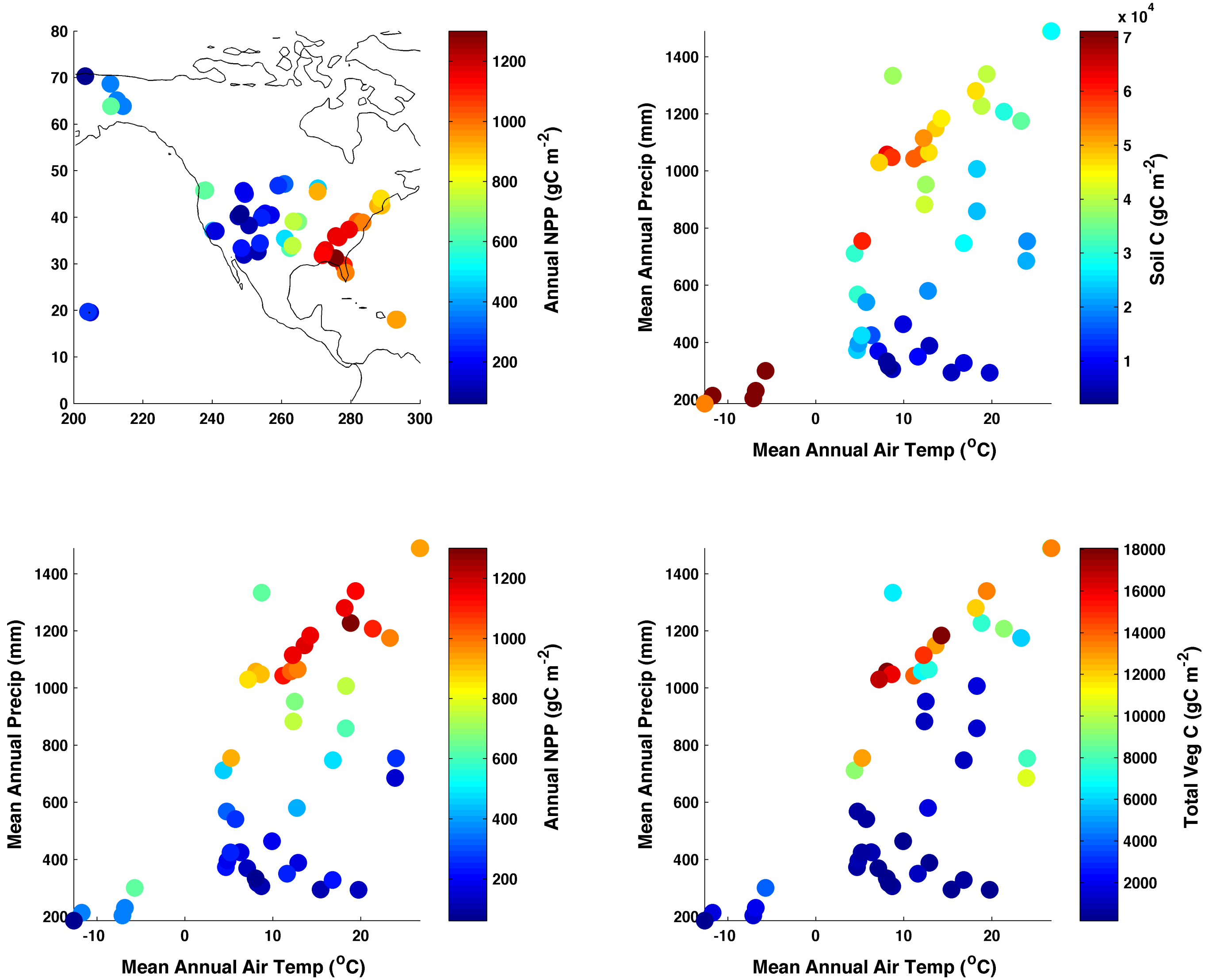
SMAP SWC

OCO-2 Fluorescence

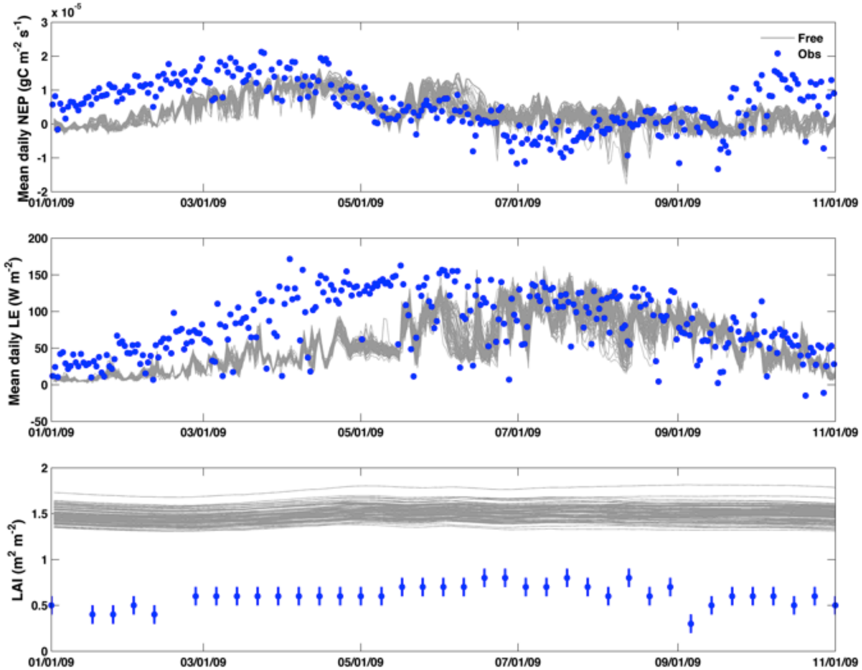


ECOSTRESS & GEDI

NEON in CLM-space

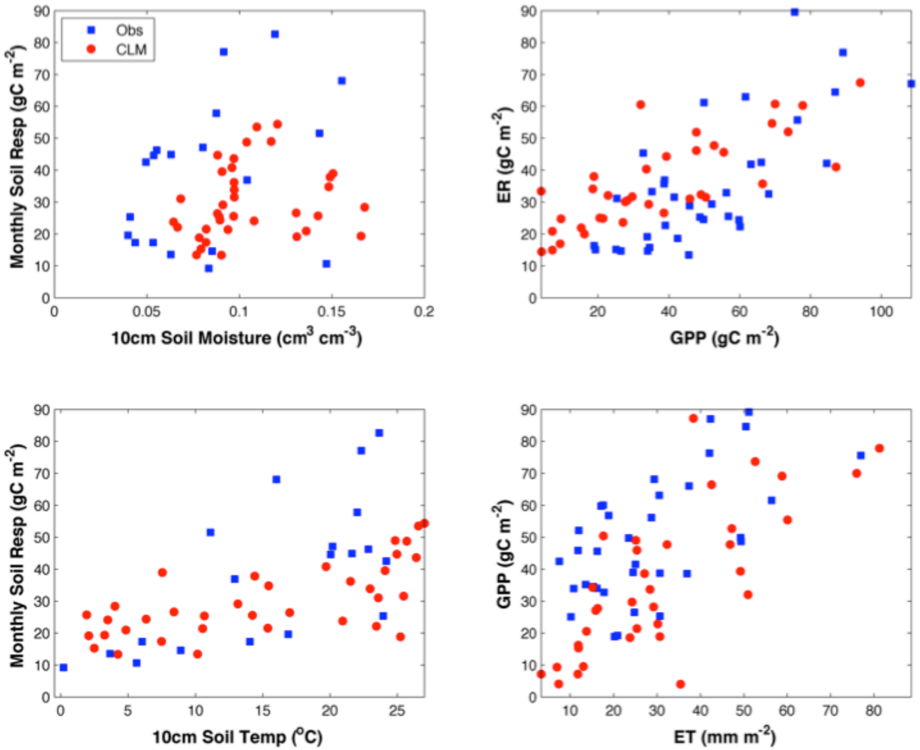


Alternative approaches

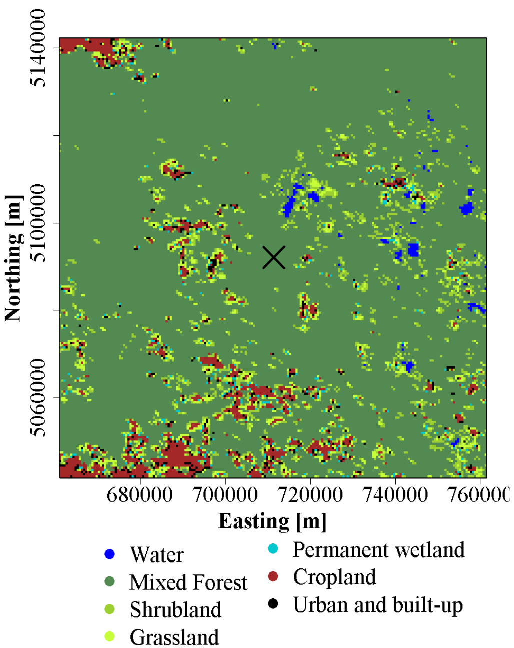


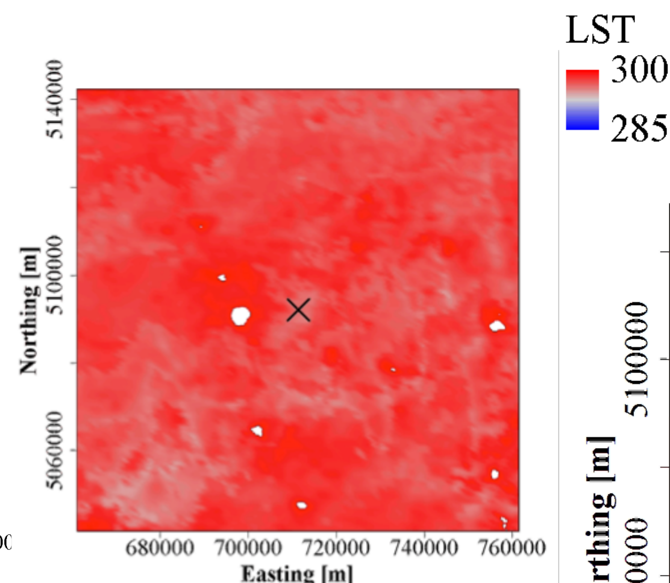
Site level runs

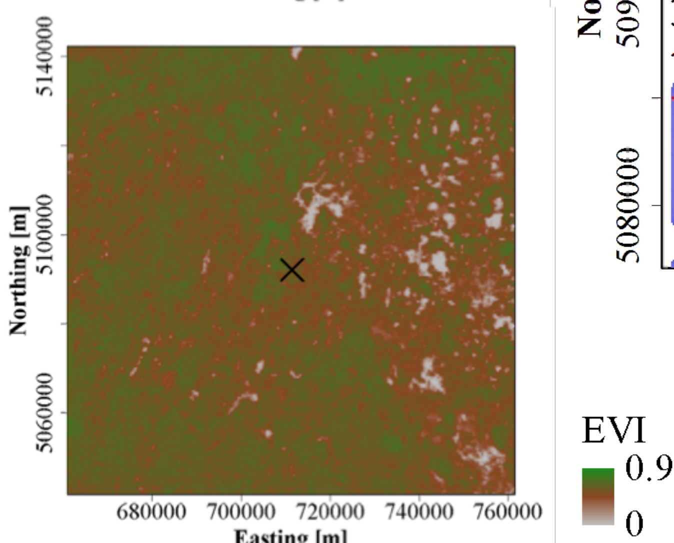
Functional Responses



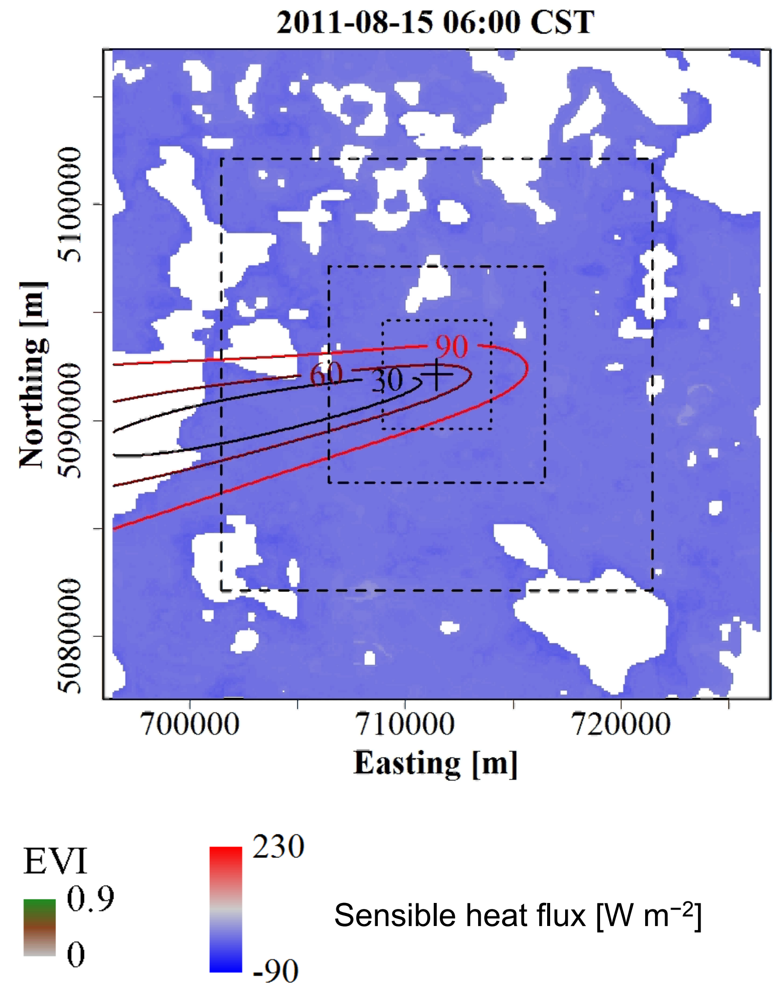
Upscale the observations







Flux grids

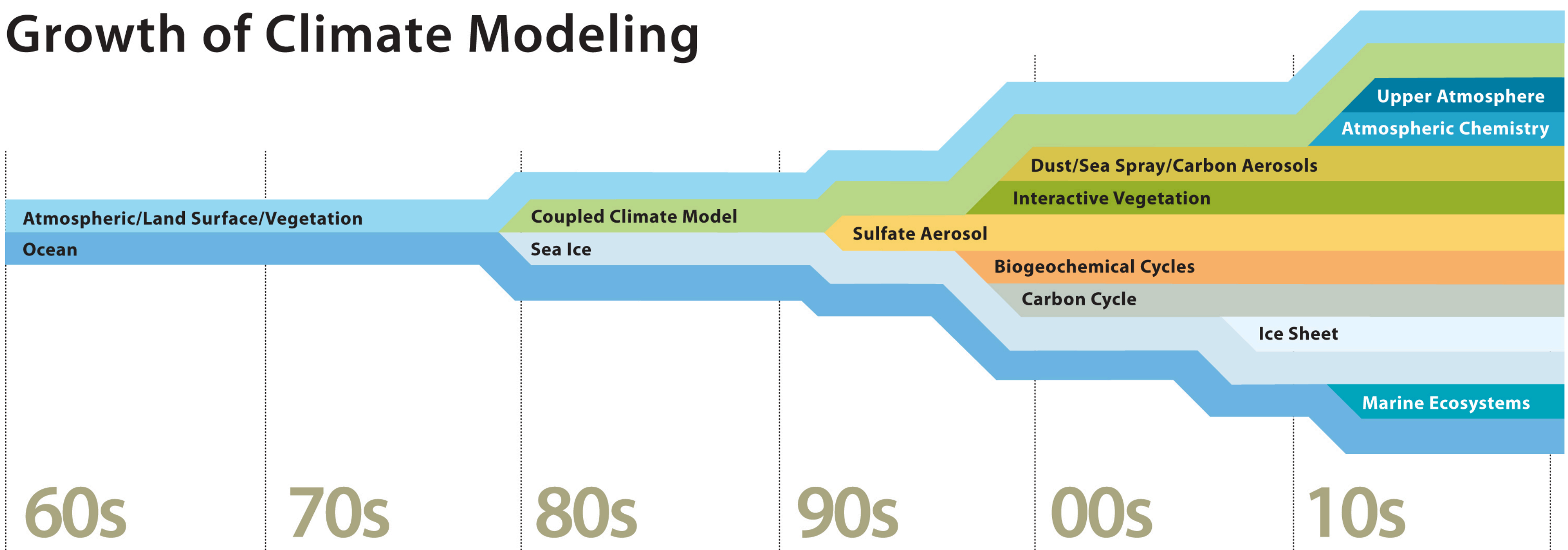


Biophysical properties used with ERF

Credit: Stefan Metzger

MODEL

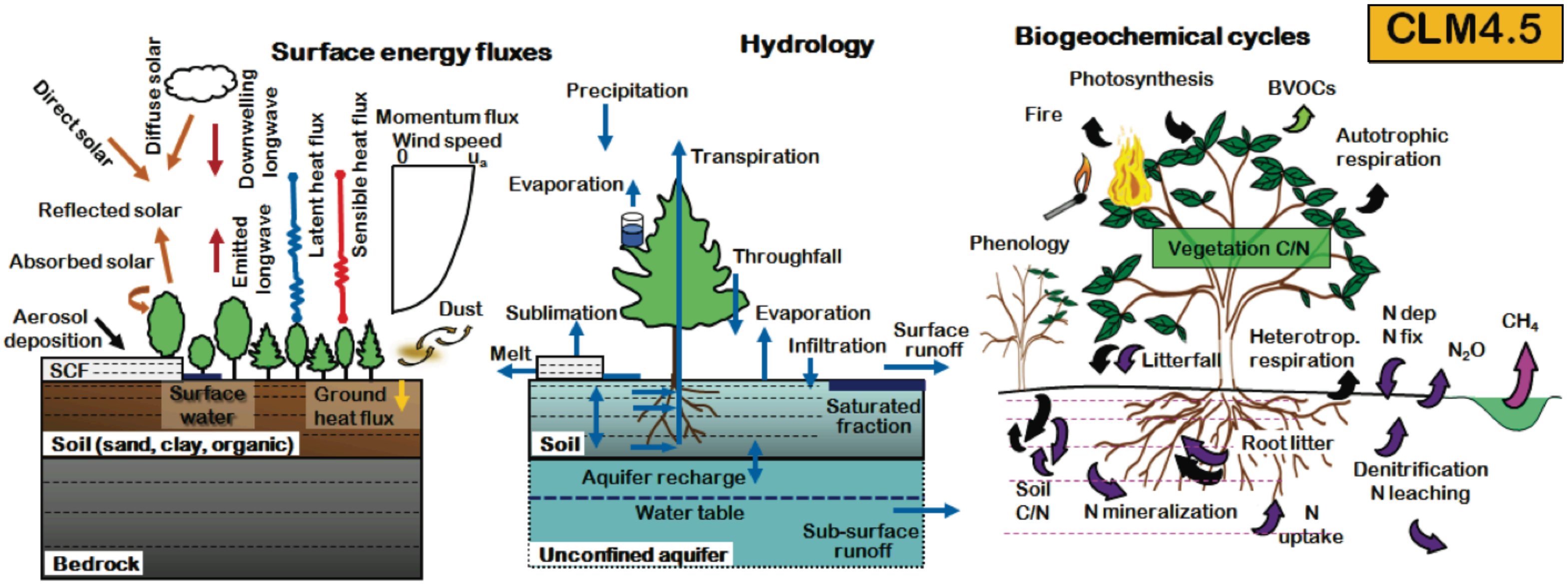
The evolution of Earth System Models

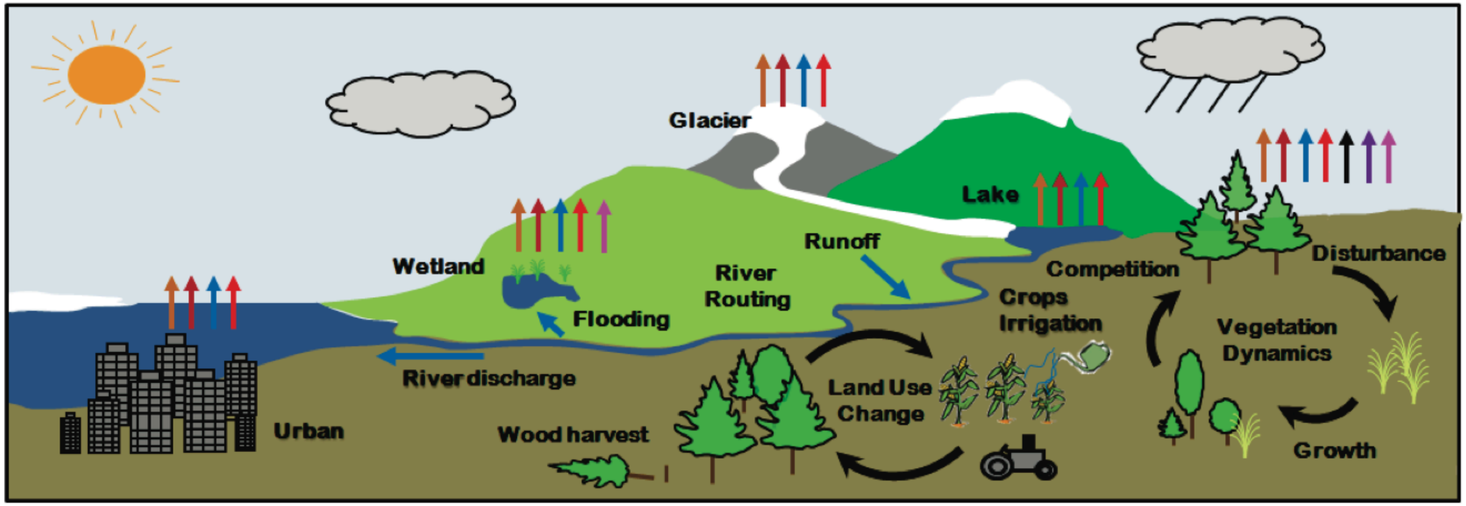


The role of the Community Land Model

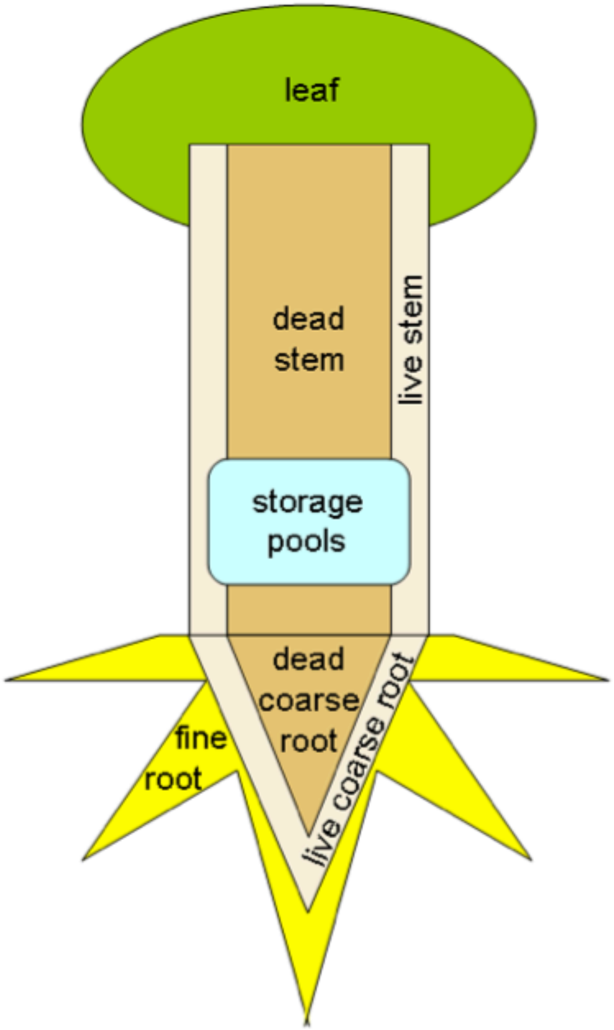
- Provide energy, water, and momentum fluxes to atmospheric model
  - Partition turbulent fluxes into latent *vs*. sensible heat
  - Determine absorbed solar radiation, surface albedo
- Runoff to ocean model
  - Riverine transport of water (and sediment, carbon, and nutrients)
- Trace gas and particle exchange to atmospheric model
  - CO_2 fluxes to atmosphere
  - CH_4, N_2O
  - Dust emissions
  - Biogenic Volatile Organic Compound emissions

The Community Land Model





Carbon and nitrogen pools



C an N pools for each tissue (structural pools)
- Leaf
- Stem (live and dead)
- Coarse root (live and dead)
- Fine root

Each structural pool has two corresponding storage pools
- Long-term storage (> 1yr)
- Short-term storage (< 1yr)

Additional pools
- Growth respiration storage (C)
- Maintenance respiration reserve (C)
- Retranslocated nitrogen

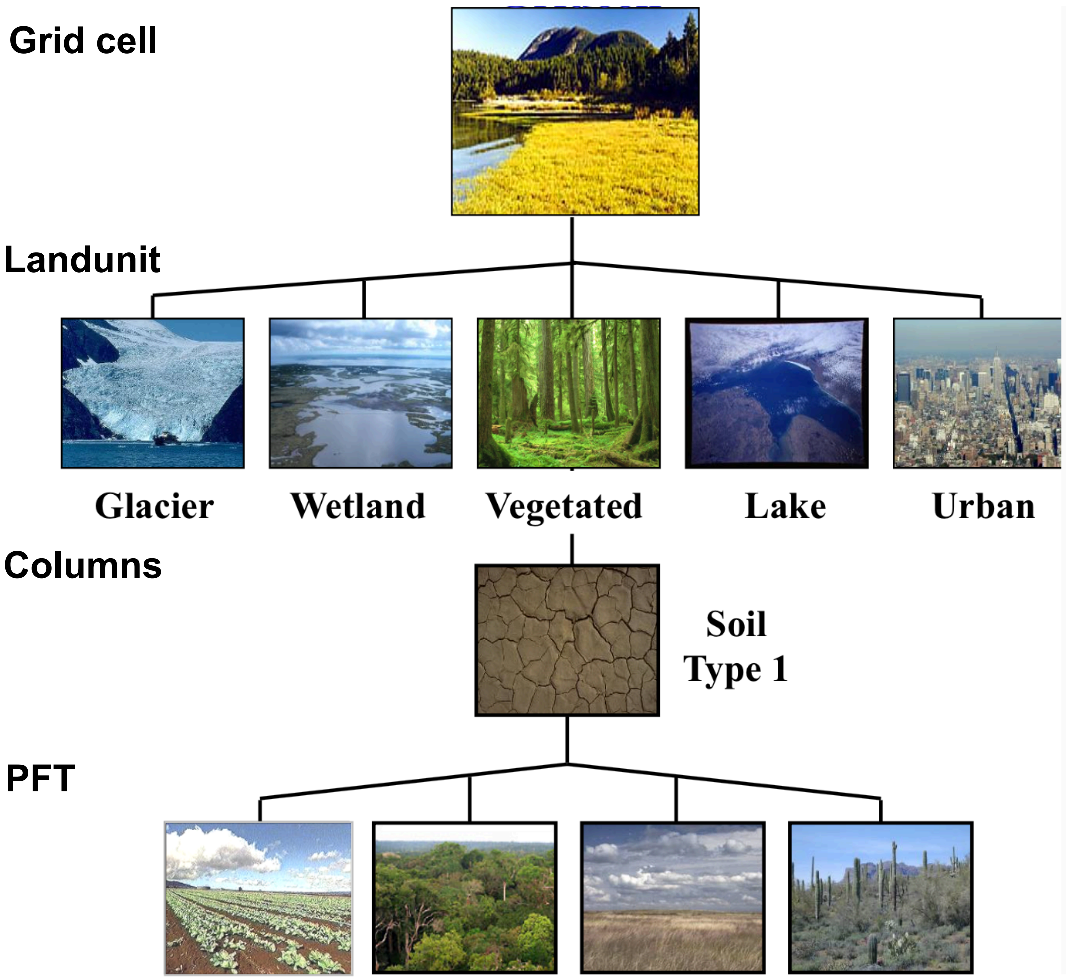
Total number of pools

Carbon: 6 + 12 + 2 = 20

Nitrogen: 6 + 12 + 1 = 19

Oleson *et al*. 2010

Subgrid tiling structure and Plant functional types



1 Bare ground
2 Needleleaf Evergreen, Temperate
3 Needleleaf Evergreen, Boreal
4 Needleleaf Deciduous, Boreal
5 Broadleaf Evergreen, Tropical
6 Broadleaf Evergreen, Temperate
7 Broadleaf Deciduous, Tropical
8 Broadleaf Deciduous, Temperate
9 Broadleaf Deciduous, Boreal
10 Broadleaf Evergreen Shrub, Temperate
11 Broadleaf Deciduous Shrub, Temperate
12 Broadleaf Deciduous Shrub, Boreal
13 C3 Arctic Grasses
14 C3 non-Arctic Grasses
15 C4 Grass
16 Crop

Specific subgrid units don't necessarily have location information

Specific observations have location information but don't normally have subgrid unit information

Main components of CLM

1. Soil hydrology and thermodynamics model
2. Photosynthesis model
3. Carbon and nitrogen cycle model
4. Vegetation dynamics model
5. Radiation and albedo model
6. River Transport model
7. Lake model
8. Urban model
9. Volatile Organic Compound emissions model
10. Dust emissions model
11. Crop model
12. Snow model
13. Carbon and water isotopes model
14. Fire model

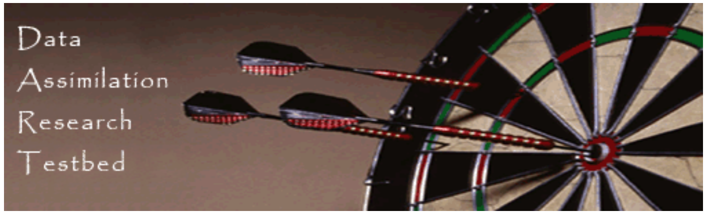
Main components of CLM

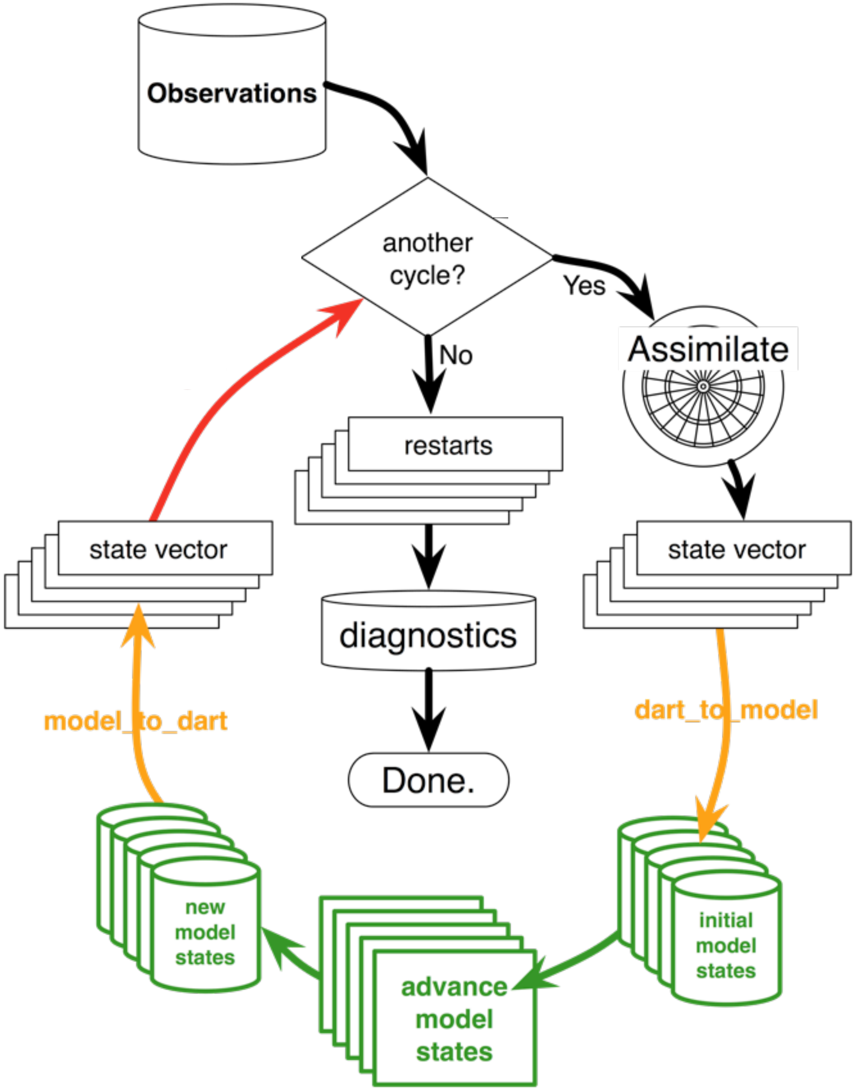
1. Soil hydrology and thermodynamics model
2. Photosynthesis model
3. Carbon and nitrogen cycle model
4. Vegetation dynamics model
5. Radiation and albedo model
6. River Transport model
7. Lake model
8. Urban model
9. Volatile Organic Compound emissions model
10. Dust emissions model
11. Crop model
12. Snow model
13. Carbon and water isotopes model
14. Fire model

DATA ASSIMILATION RESEARCH TESTBED

Data Assimilation Research Testbed (DART)

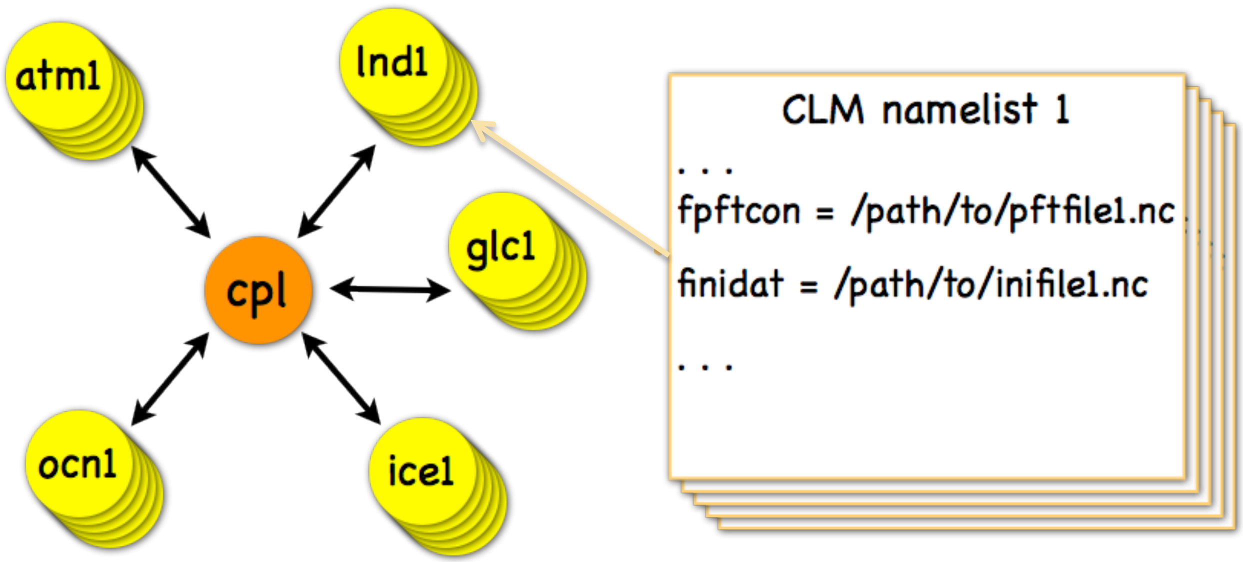
- DART is a community facility for ensemble DA
- Uses a variety of flavors of filters
  - Ensemble Adjustment Kalman Filter
- Many enhancements to basic filtering algorithms
  - Adaptive inflation
  - Localization
- Uses new multi-instance capability within CESM





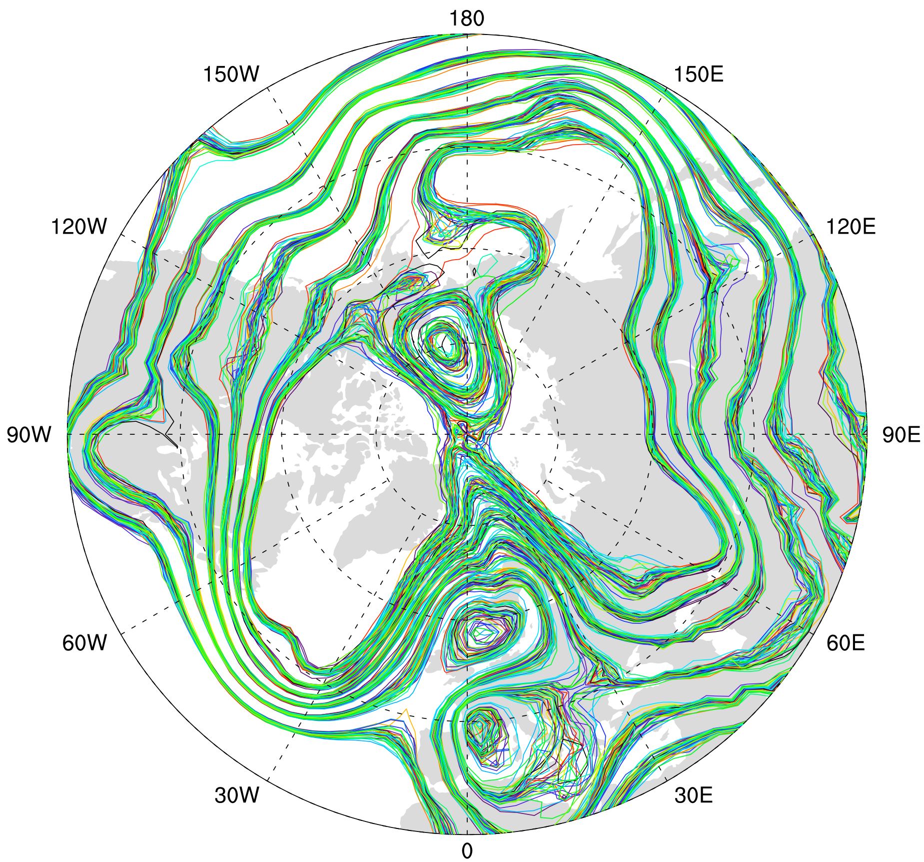
Multi-instance CESM code

- A multi-instance version of CESM has been developed that more easily facilitates ensemble-based DA
- For example, multiple land models can be driven by multiple data-atmospheres in a single executable.
- This capability is available in the current CESM release.



Multi-instances of data atmospheres

- 80 member, 6 hourly reanalysis available, 1998 - 2010
- Assimilation uses 80 members of 2^o FV CAM forced by a single ocean
- O (1 million) atmospheric obs are assimilated every day
- Each CLM ensemble member is forced with a different atmospheric reanalysis member
- Generates spread in the land model



500 hPa GPH
Feb 17 2003

CLM-DART coupling

- Our goal has been to "**Do no harm**" to CLM
- DART's namelist allows you to choose what **CLM variables** get updated by the assimilation

```
&clm_vars_nml
   clm_state_variables = 'frac_sno',     'KIND_SNOWCOVER_FRAC',
                         'DZSNO',        'KIND_SNOW_THICKNESS',
                         'H2OSNO',       'KIND_SNOW_WATER',
                         'T_SOISNO',     'KIND_SOIL_TEMPERATURE',
                         'leafc',        'KIND_LEAF_CARBON' /
```

- At predetermined assimilation time step:

1. CLM stops and writes restart and history files
2. DART state vector extracted
3. Increments calculated by filter
4. Restart file updated with adjusted DART state vector
5. CLM restarts

neon National Ecological Observatory Network

Observations we can use with CLM-DART

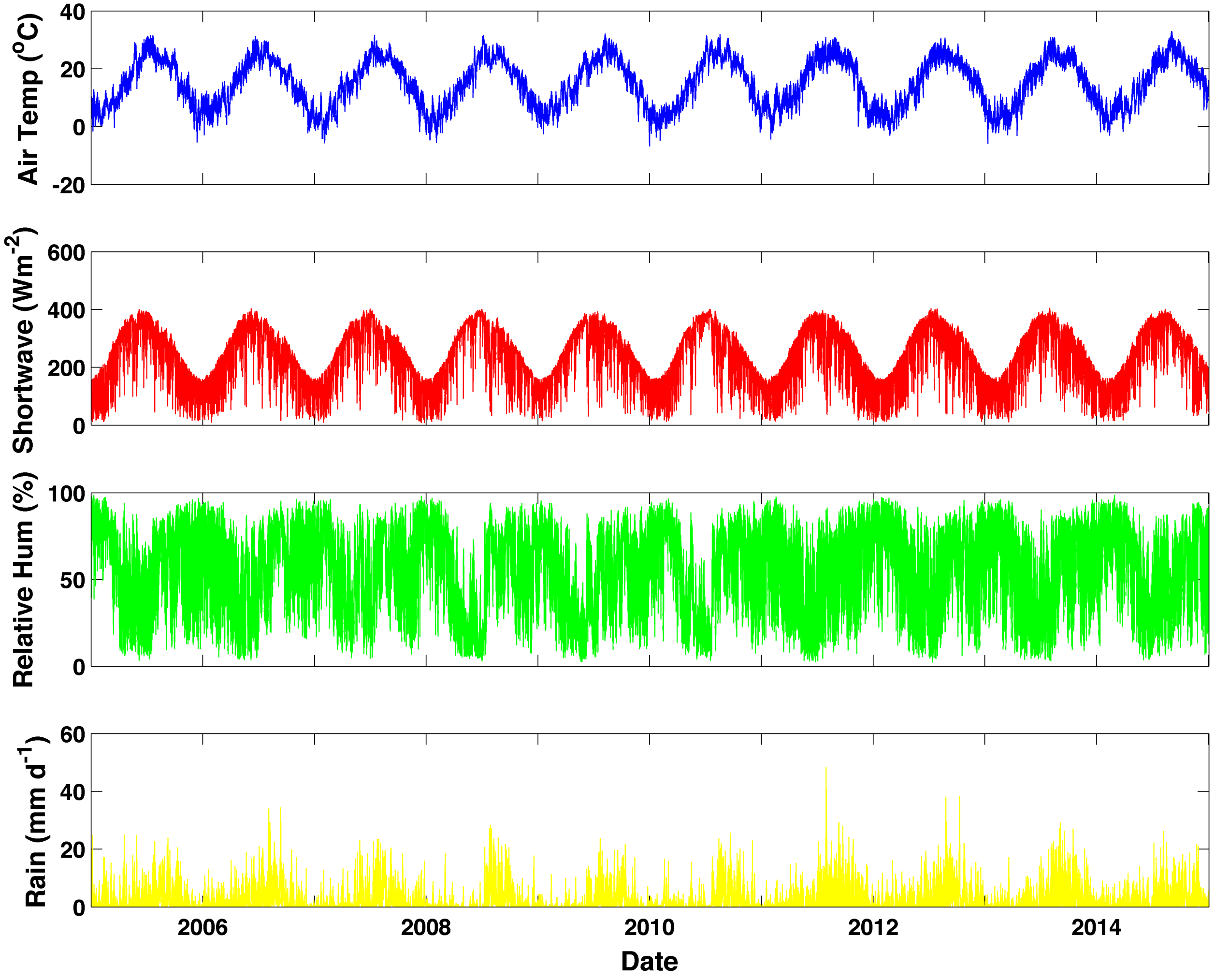
- Leaf area index
- Above ground biomass
- Canopy nitrogen
- Snow cover fraction
- Microwave brightness temperature
- Cosmic ray neutron intensities
- Total water storage anomalies (GRACE)
- Soil moisture and temperature
- Latent heat flux
- Sensible heat flux
- Carbon fluxes (NEP, GPP, ER, SR)

PERFECT MODEL EXAMPLES

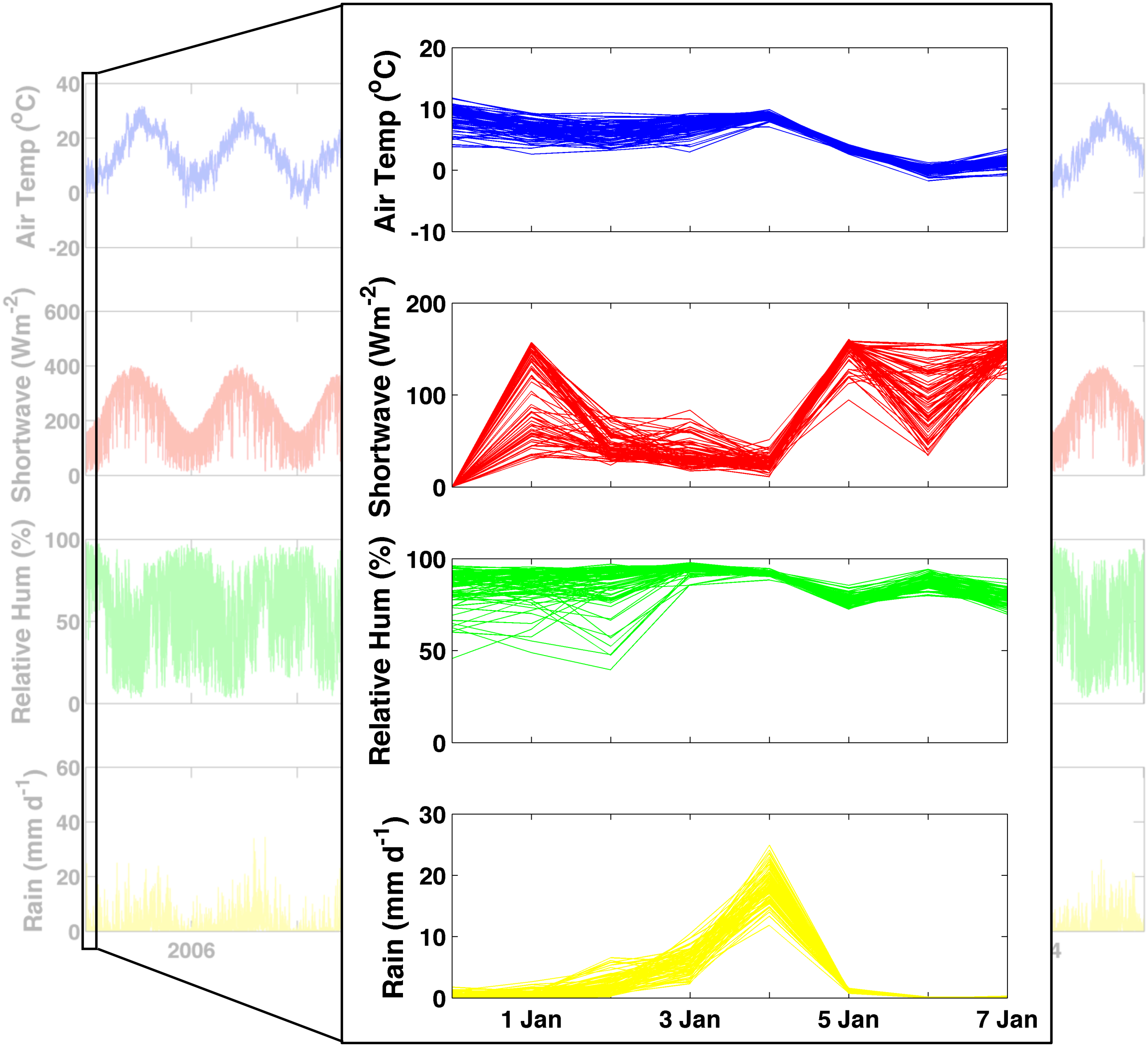
Observing System Simulation Experiment

- CLM spun up at NM Piñon-Juniper Ameriflux site
- 80 member model ensemble run forward for “decades”
- One ensemble member treated as truth
- Truth “observed” periodically with prescribed observation uncertainties
- These synthetic observations are then assimilated

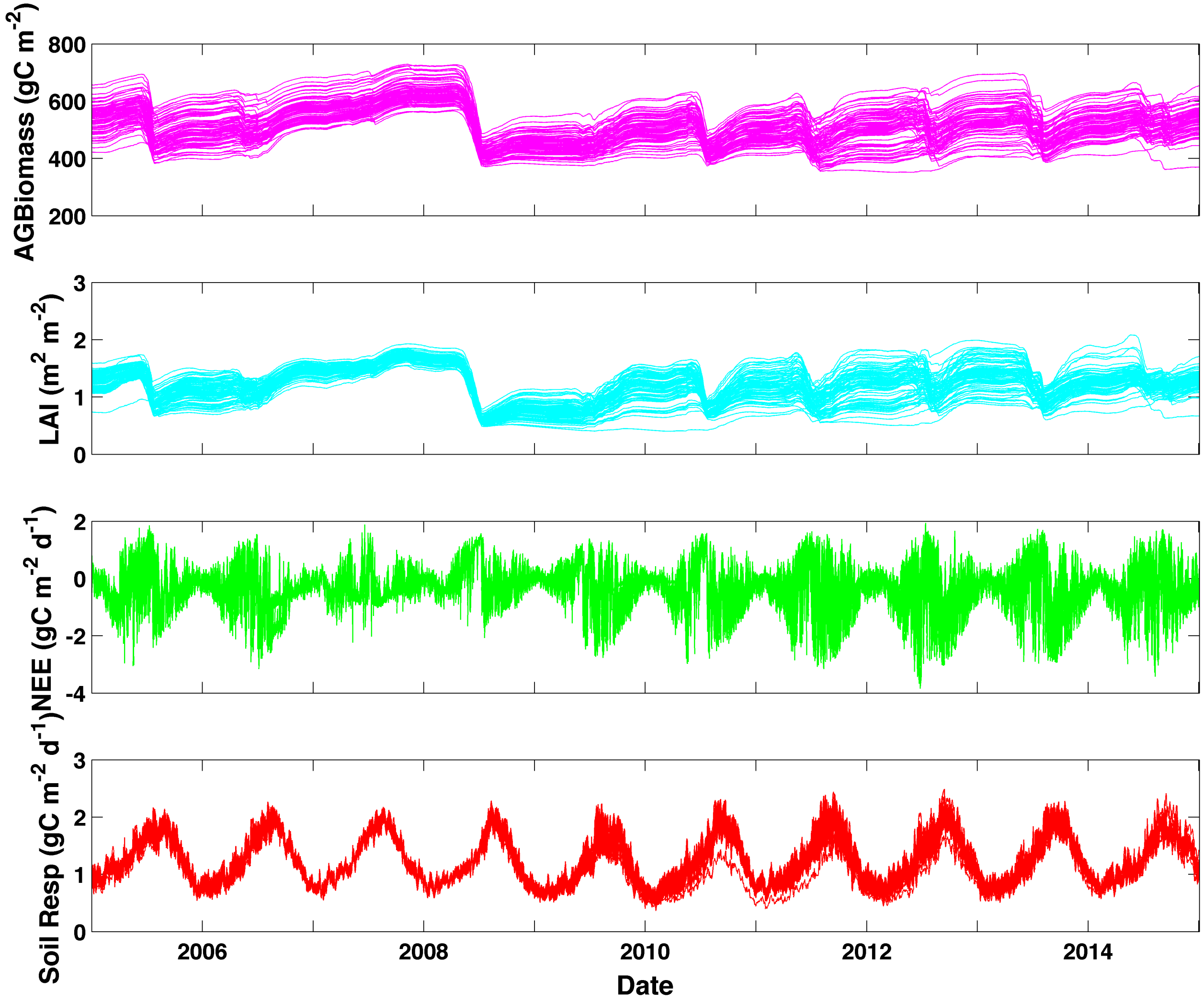
Ensemble of climate forcing



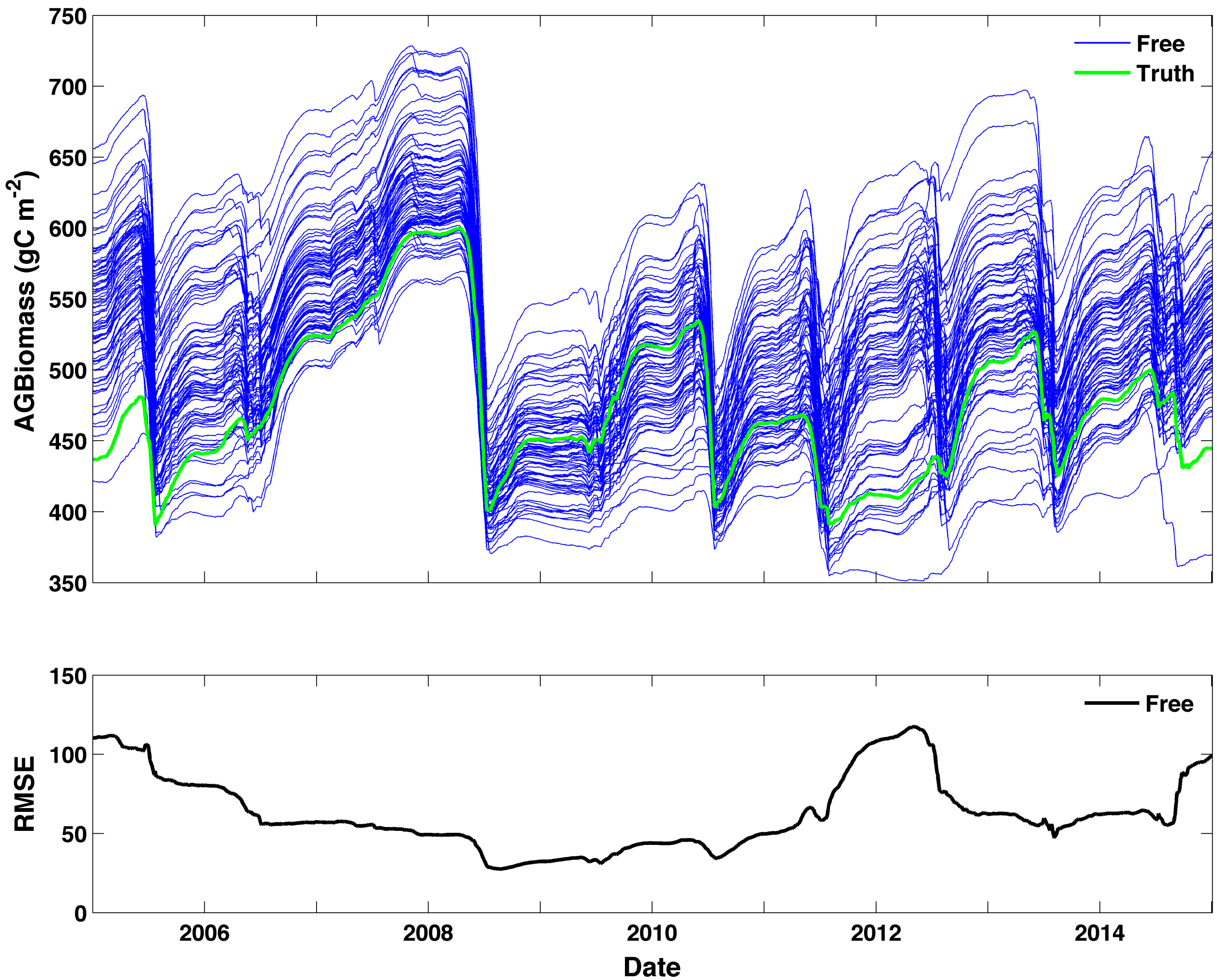
Ensemble of climate forcing



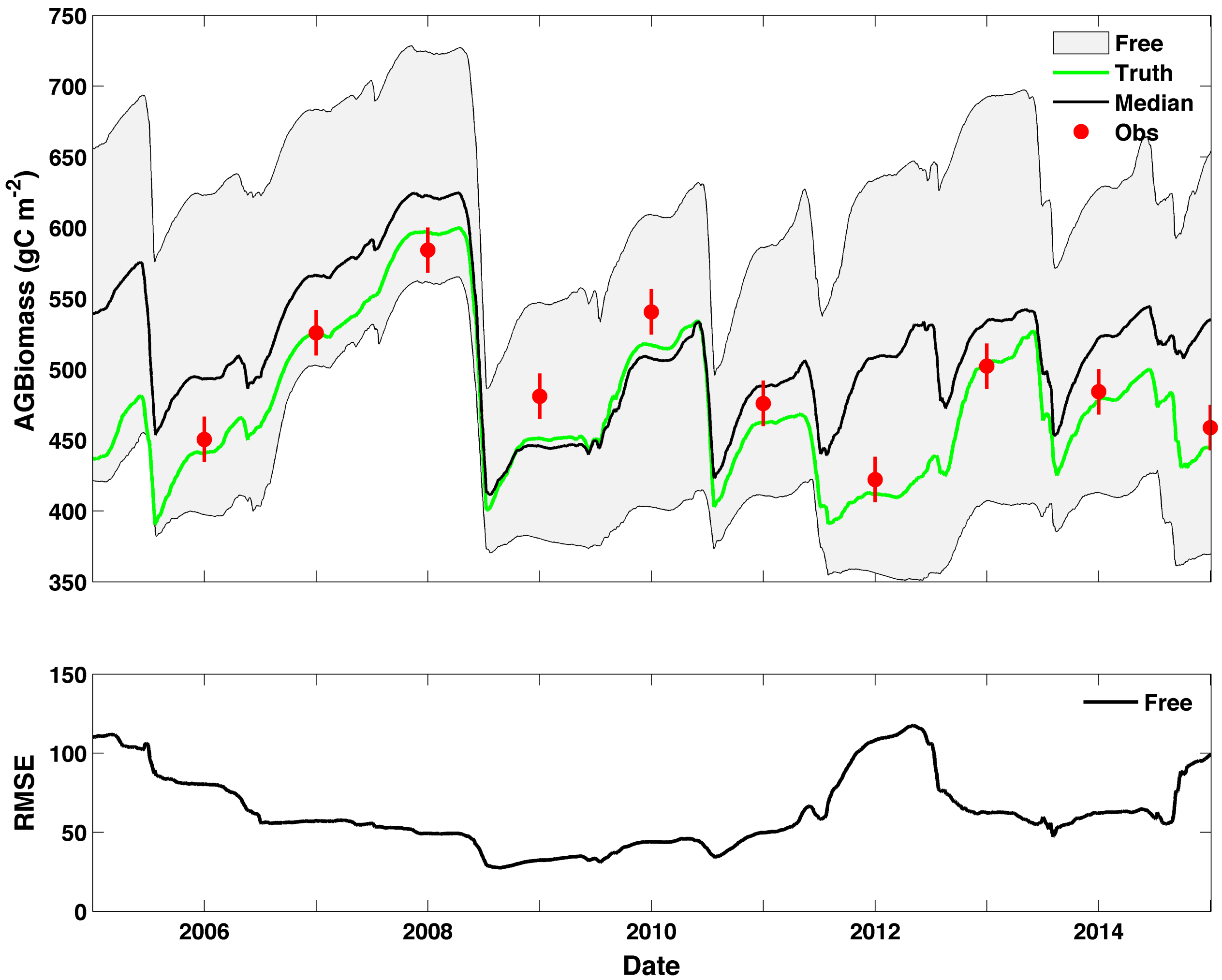
Ensemble of land surface states



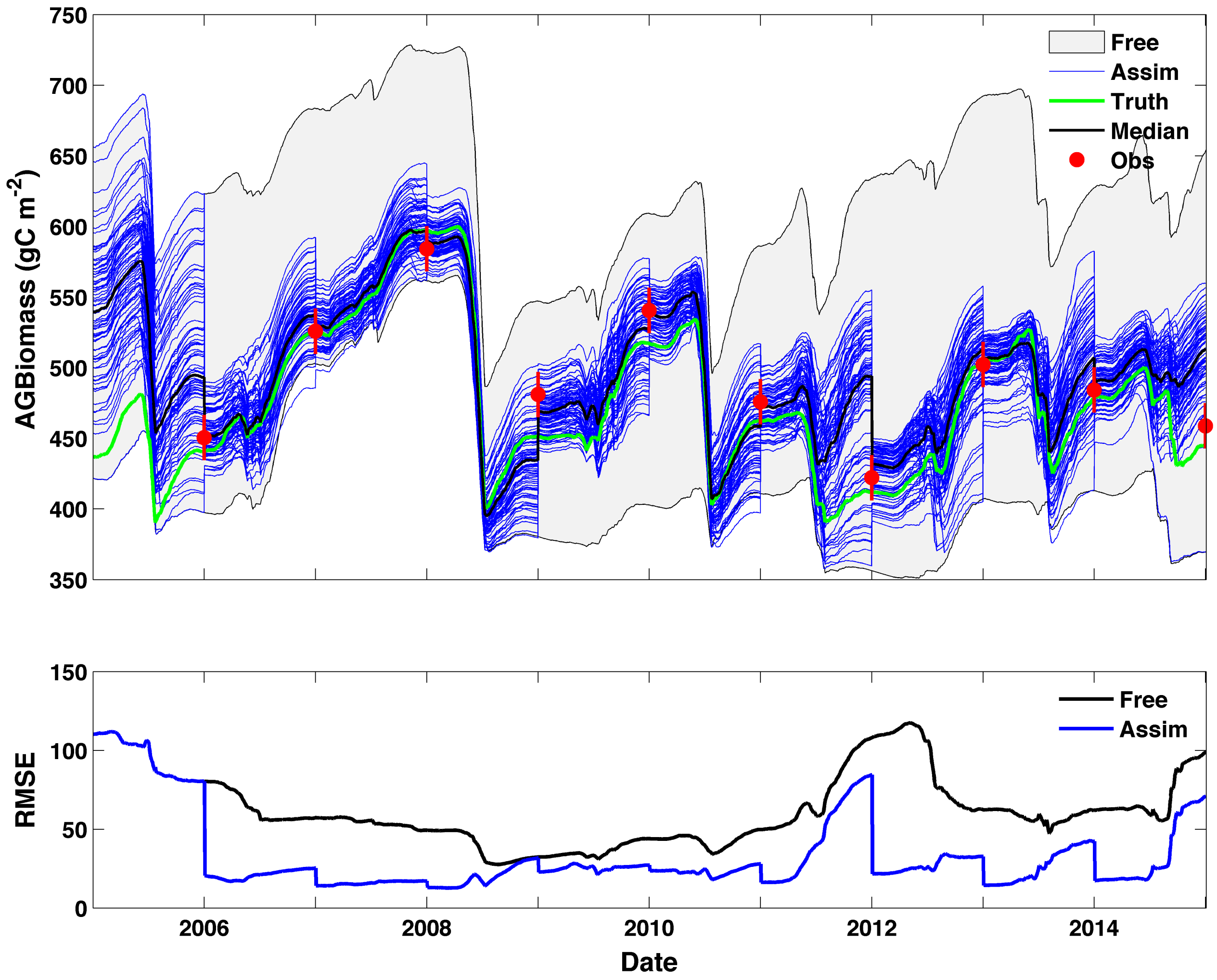
Biomass ensemble with "truth"



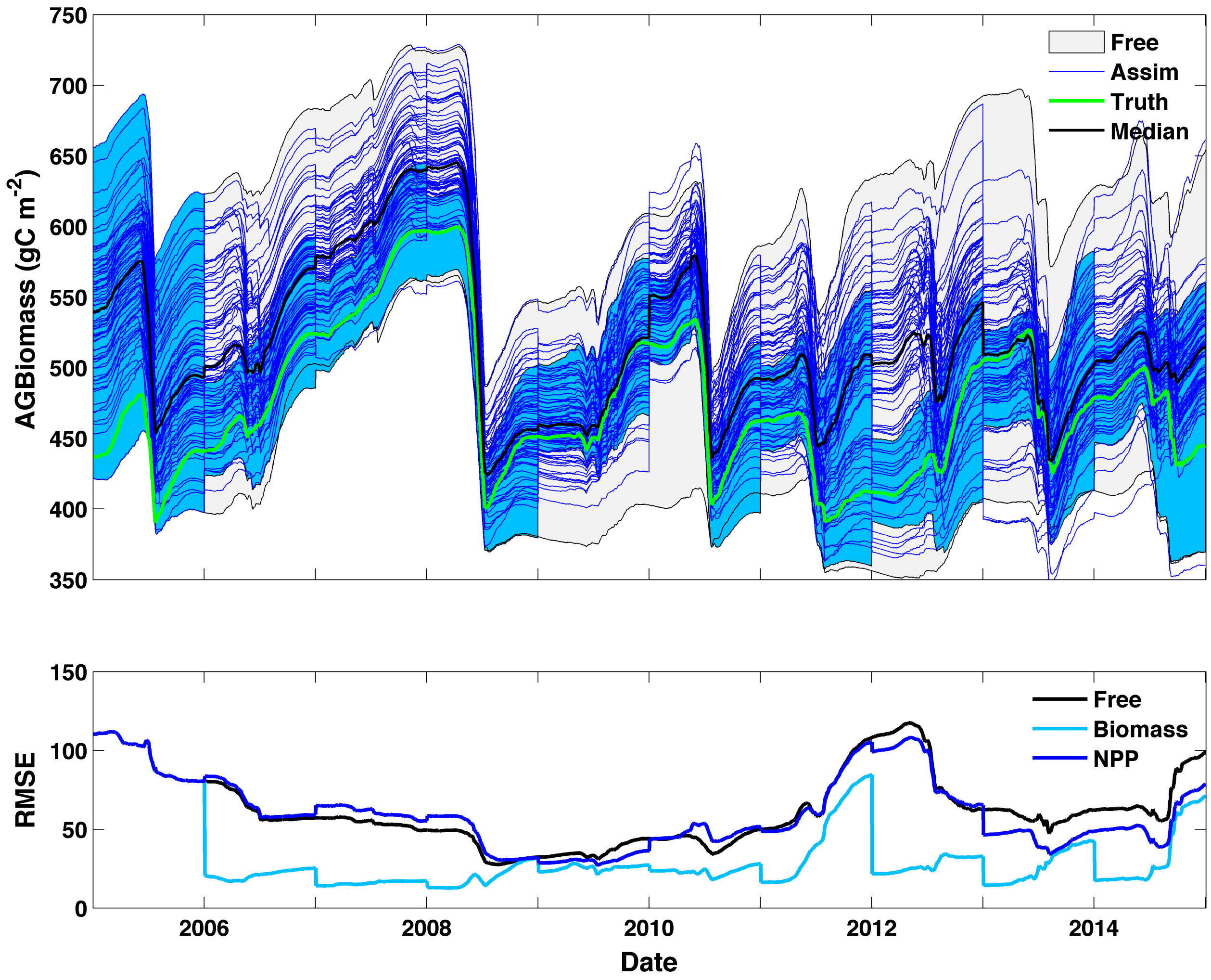
Observations around the "truth"



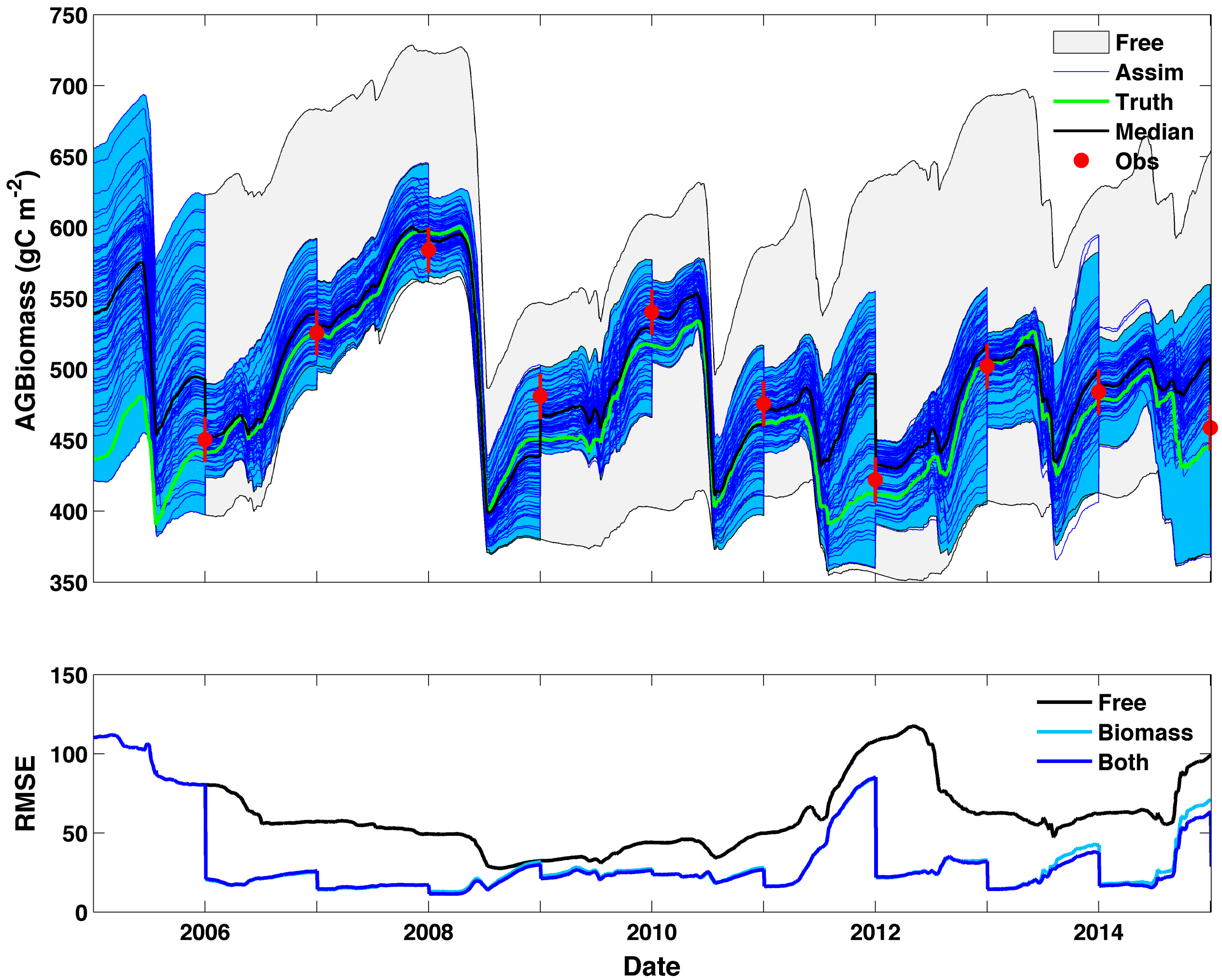
Assimilation of those synthetic observations



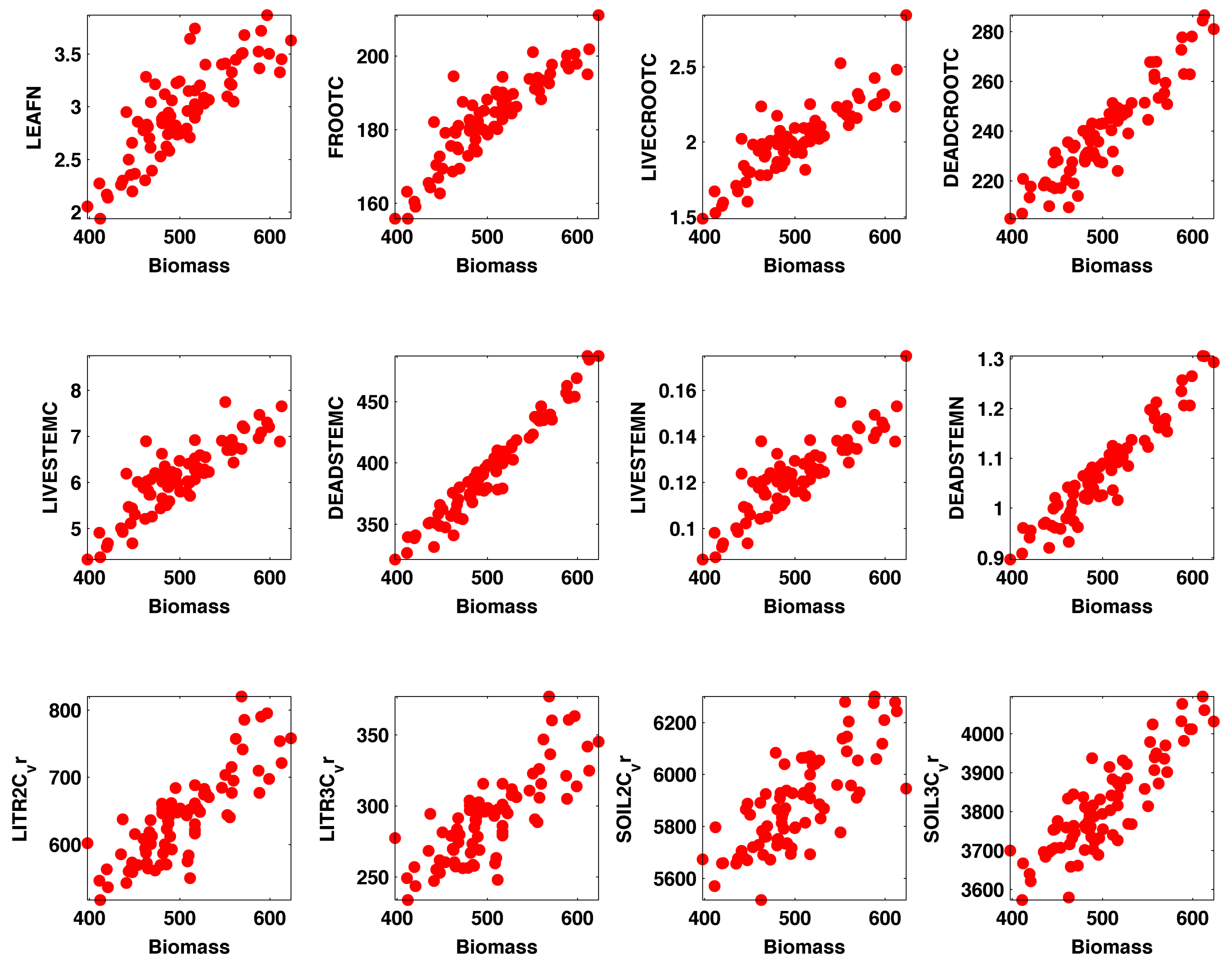
Effects of annual NPP on Biomass



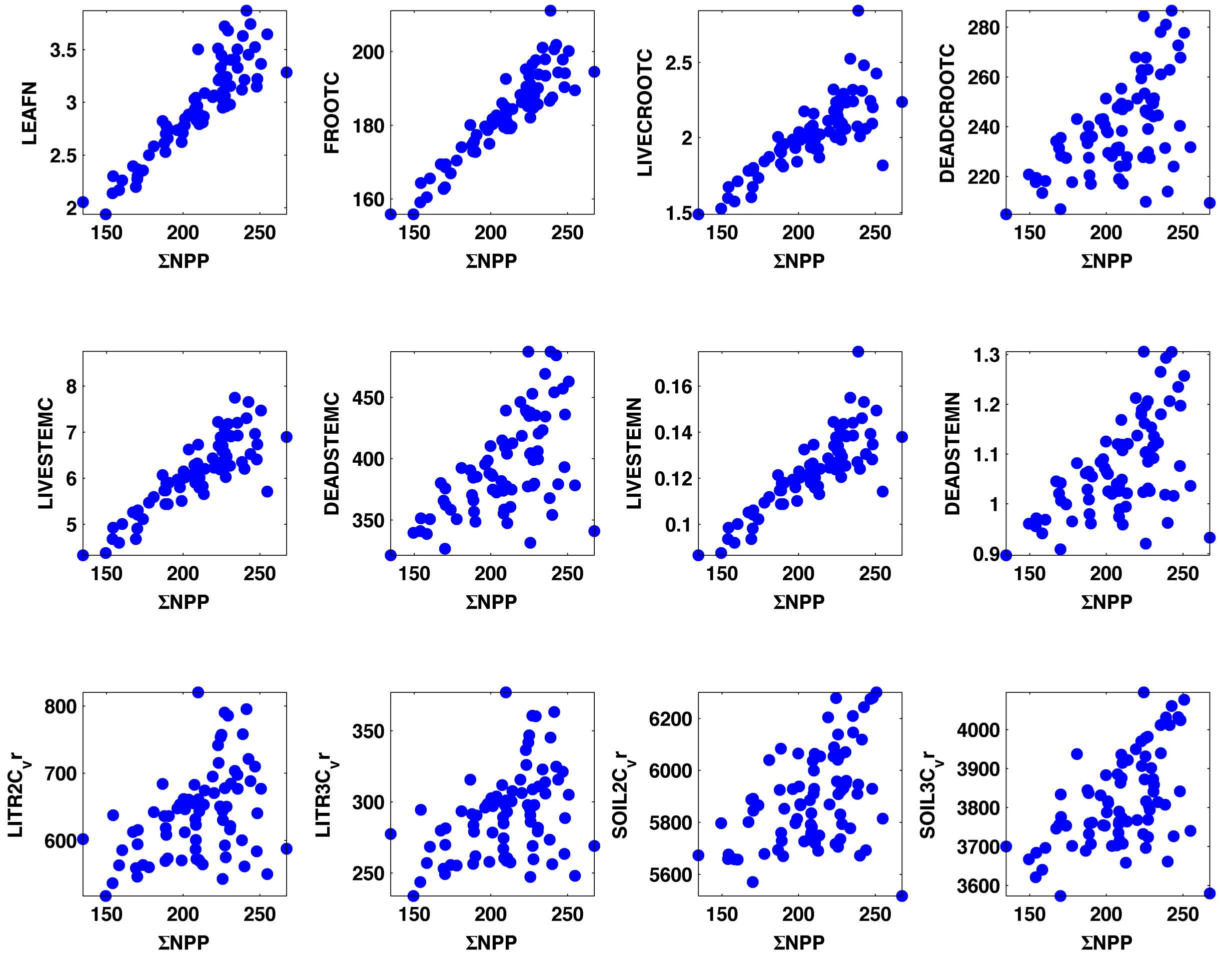
Assimilating both Biomass and annual NPP



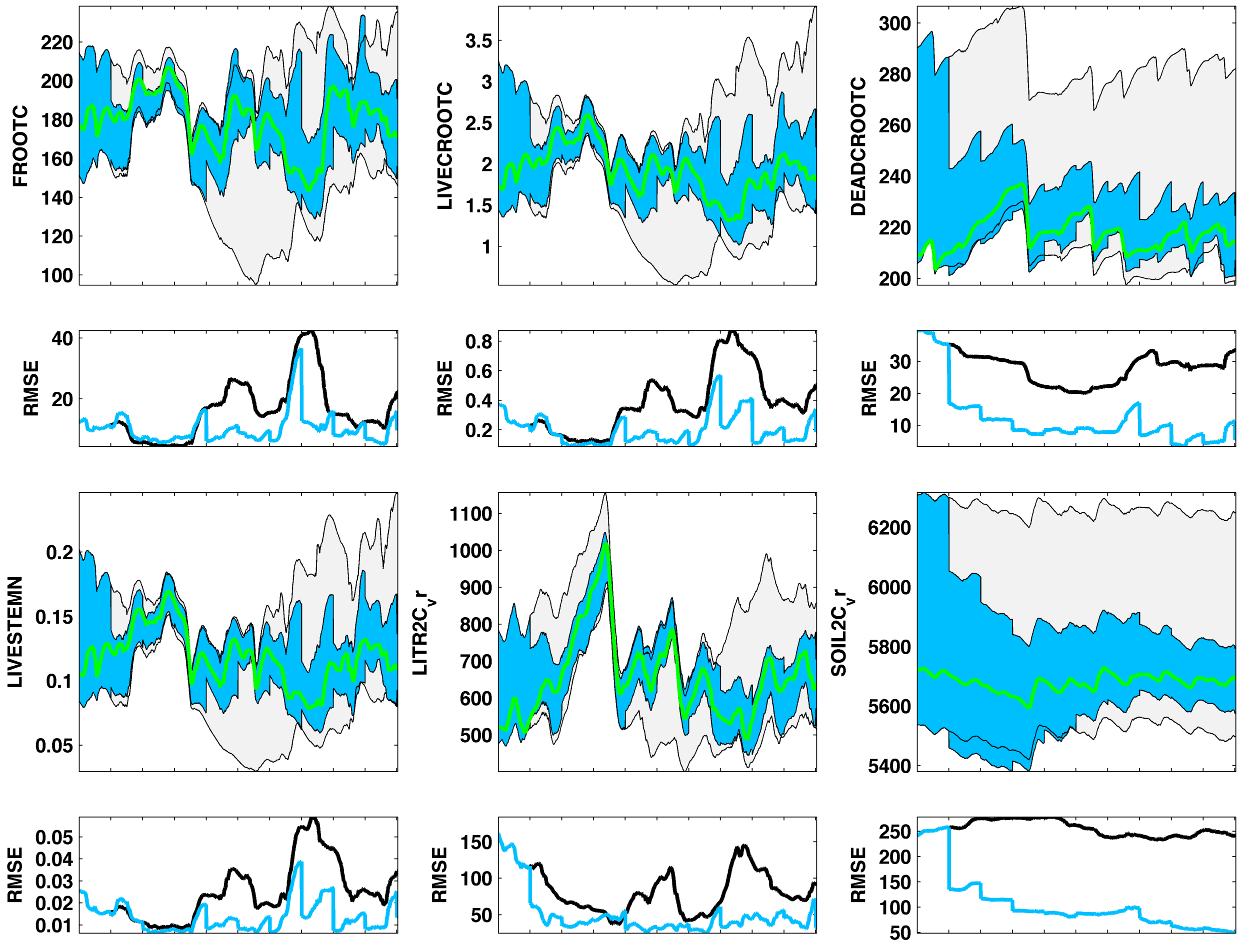
Correlations with Biomass



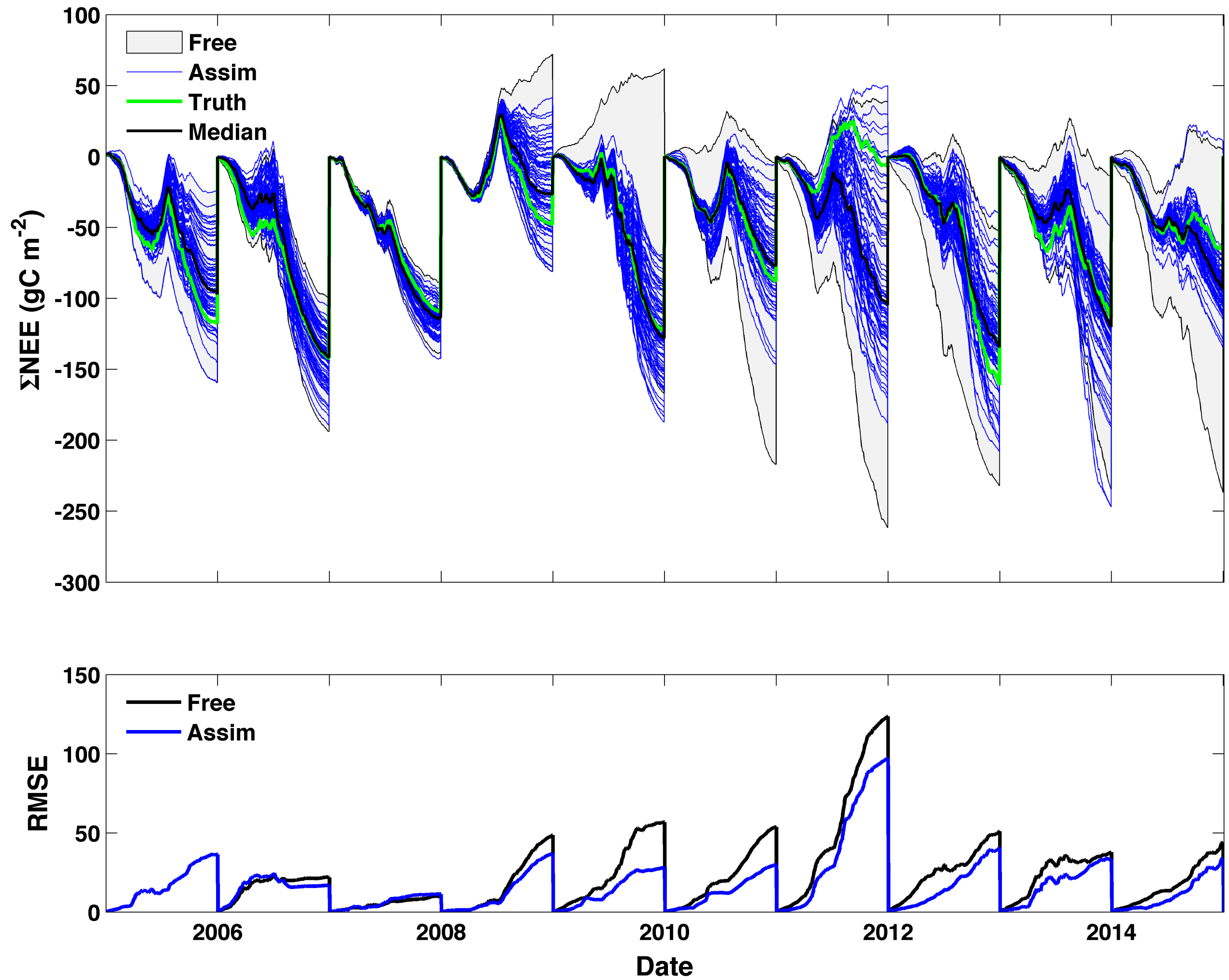
Correlations with annual NPP



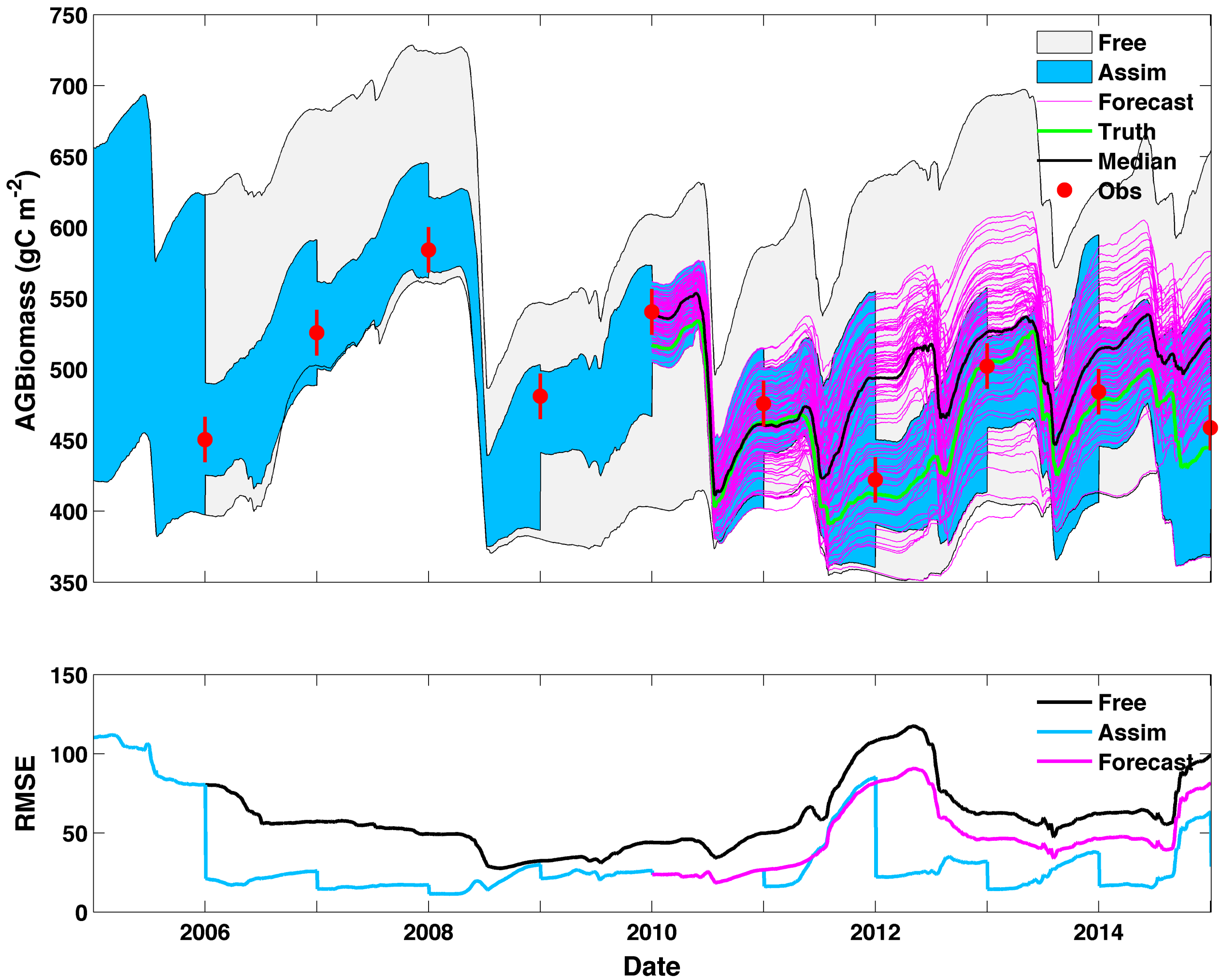
Updating unobserved states



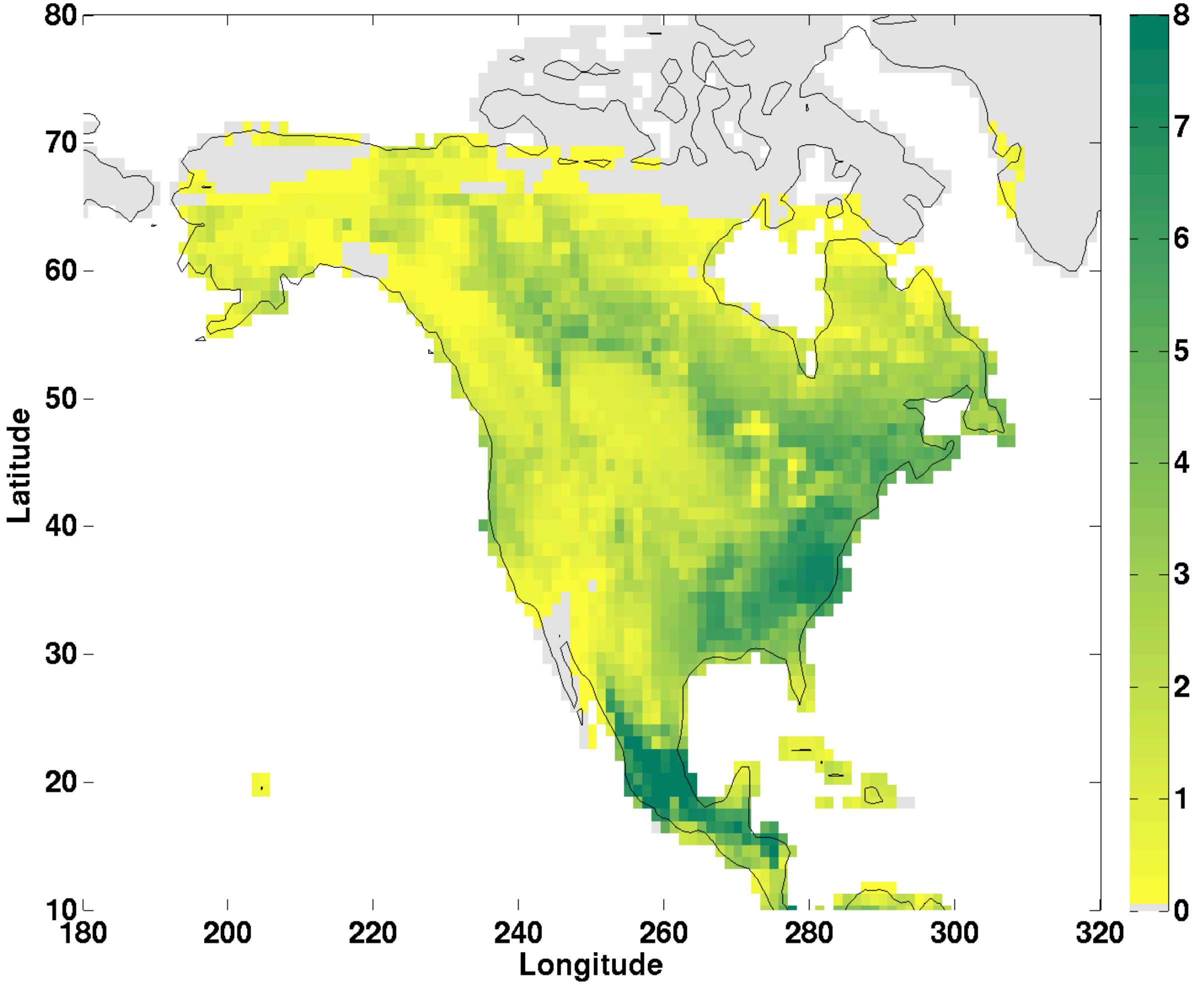
Impact of assimilation on NEE



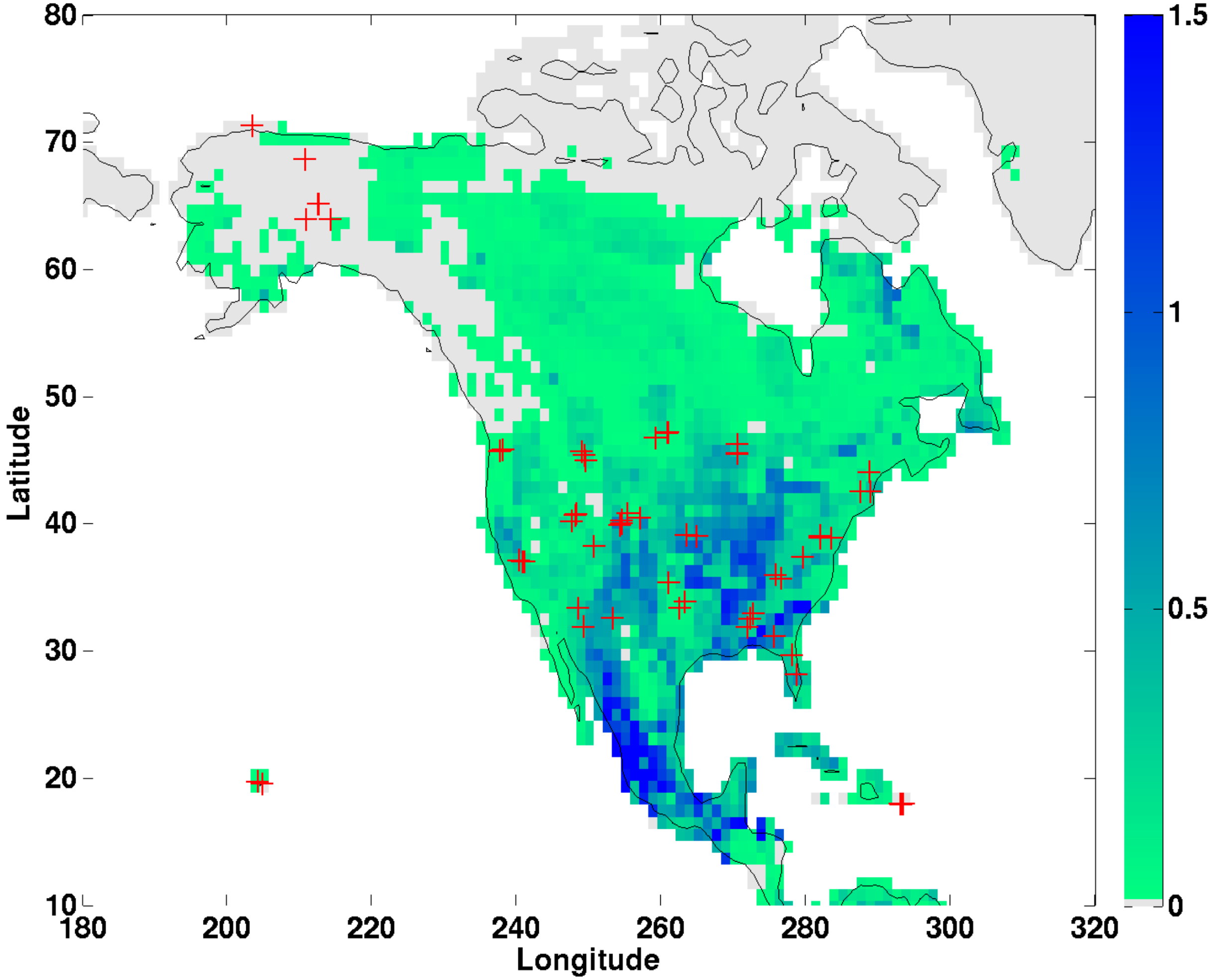
Effects on forecast



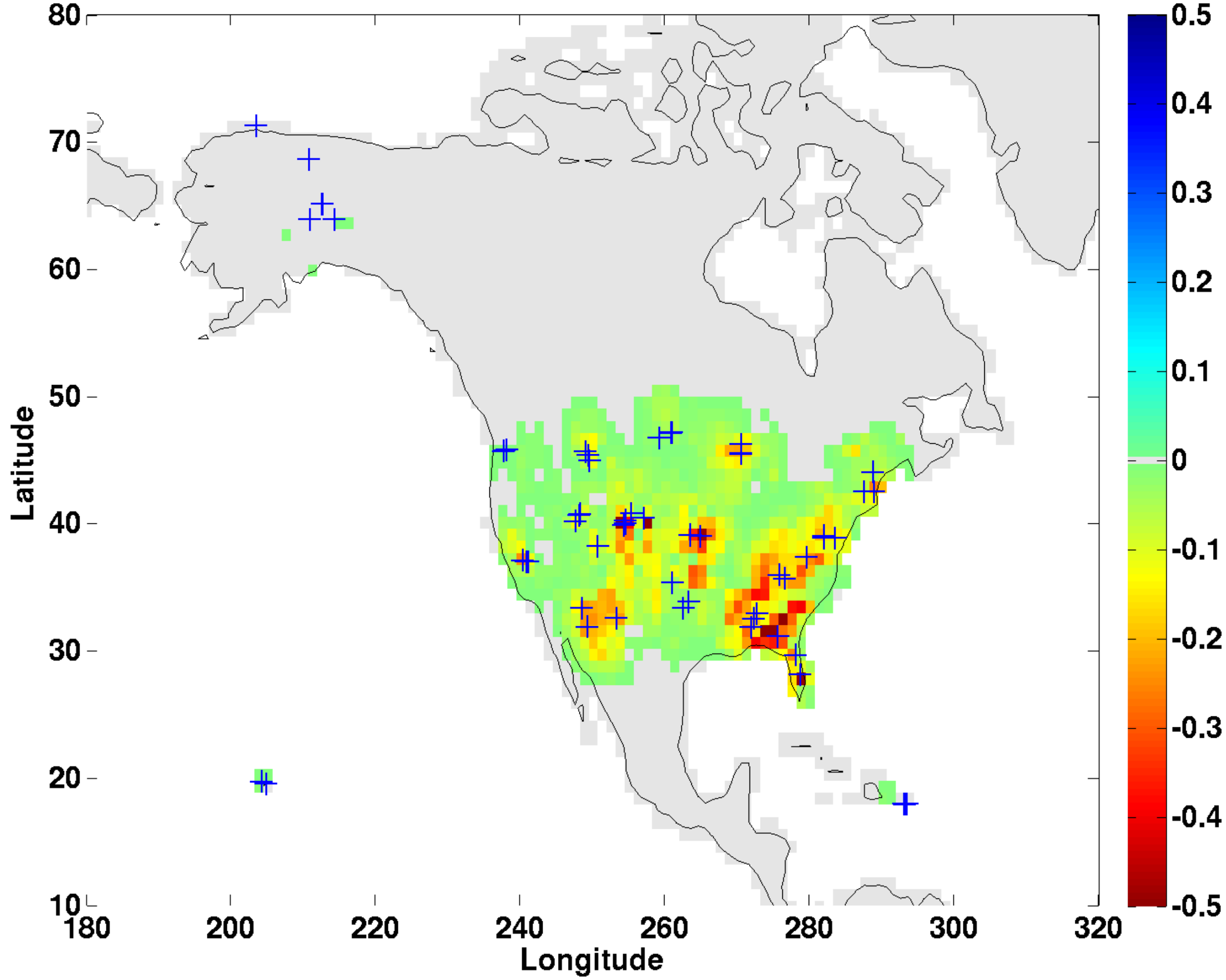
Mean LAI from 80 ensemble members



LAI spread from 80 ensemble members

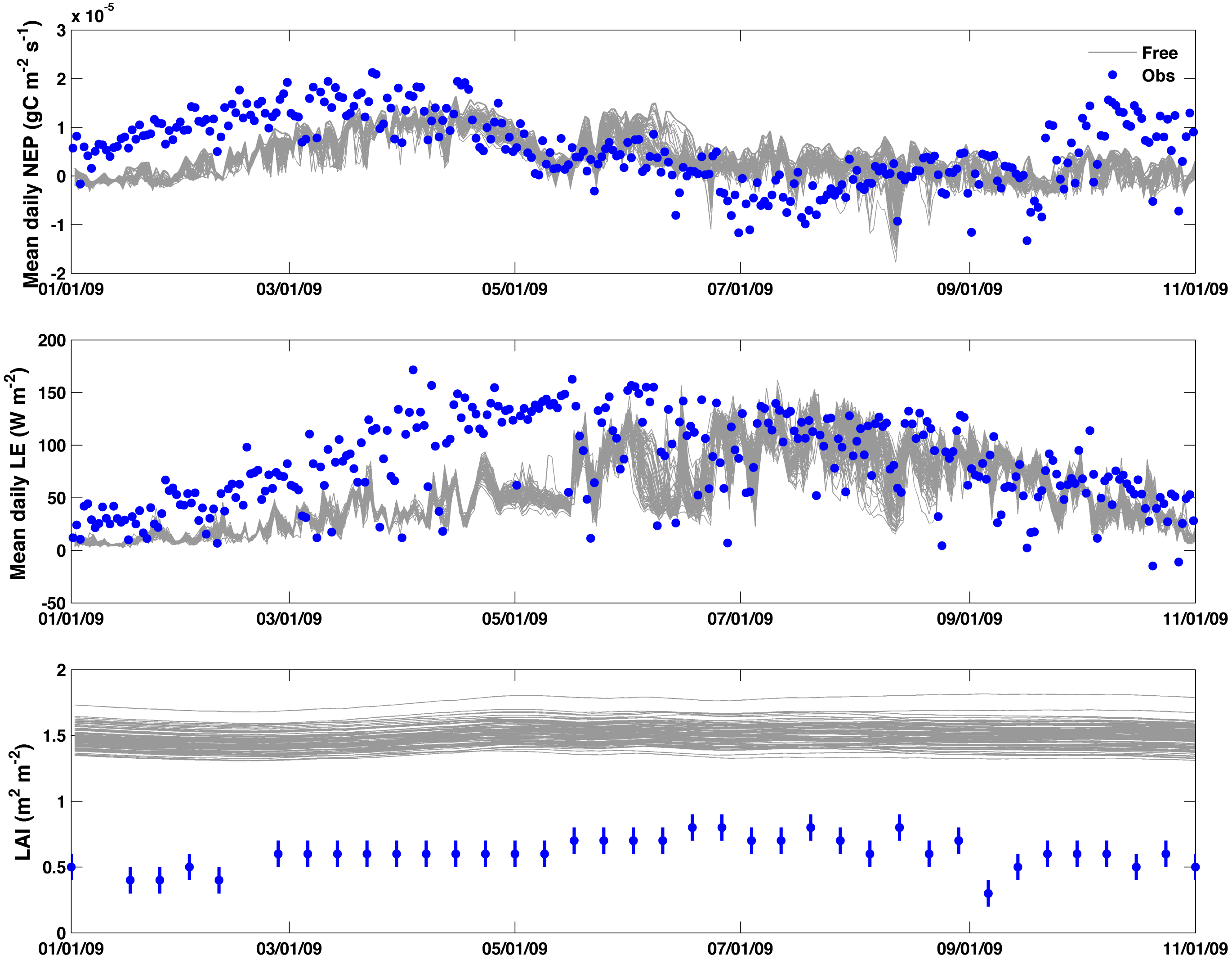


Reduction in LAI ensemble spread

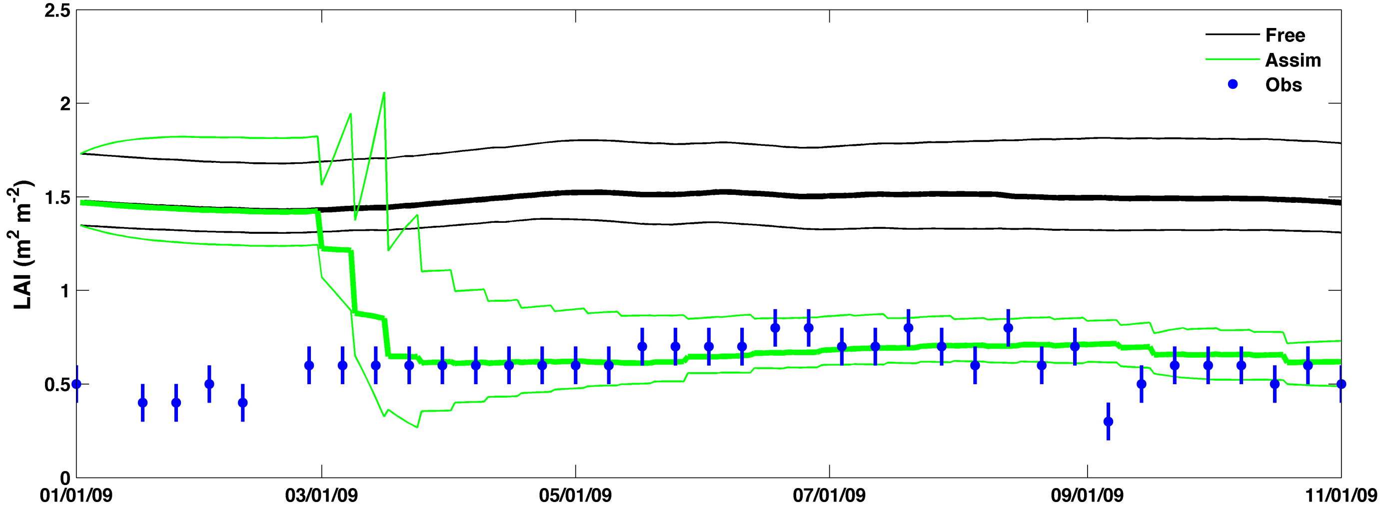


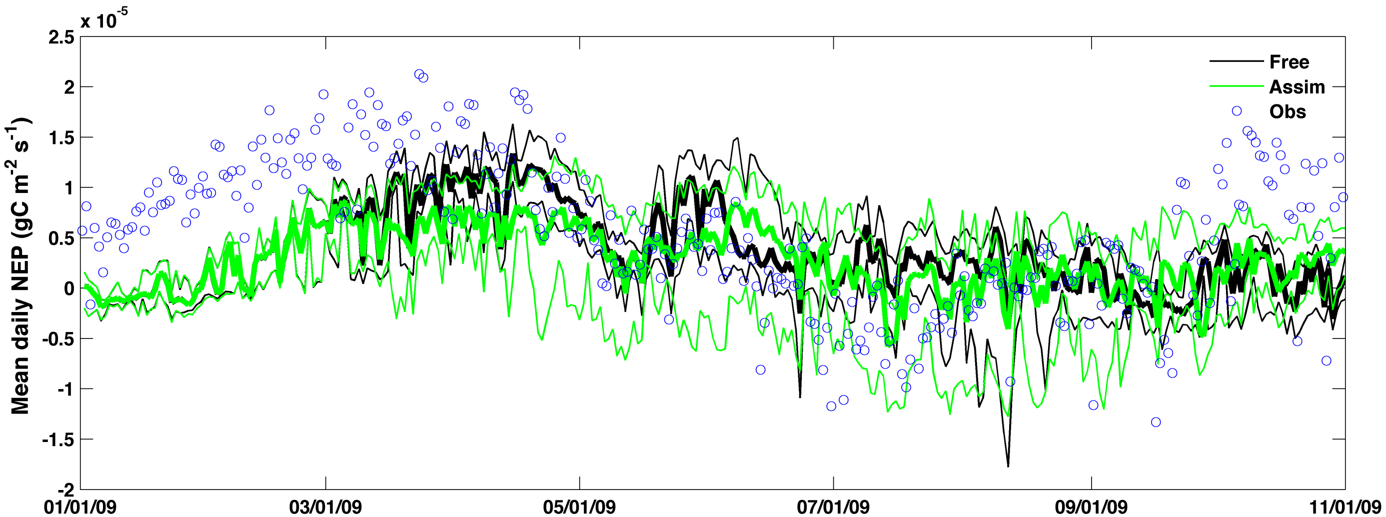
SOME REAL DATA

Ameriflux and MODIS LAI observations

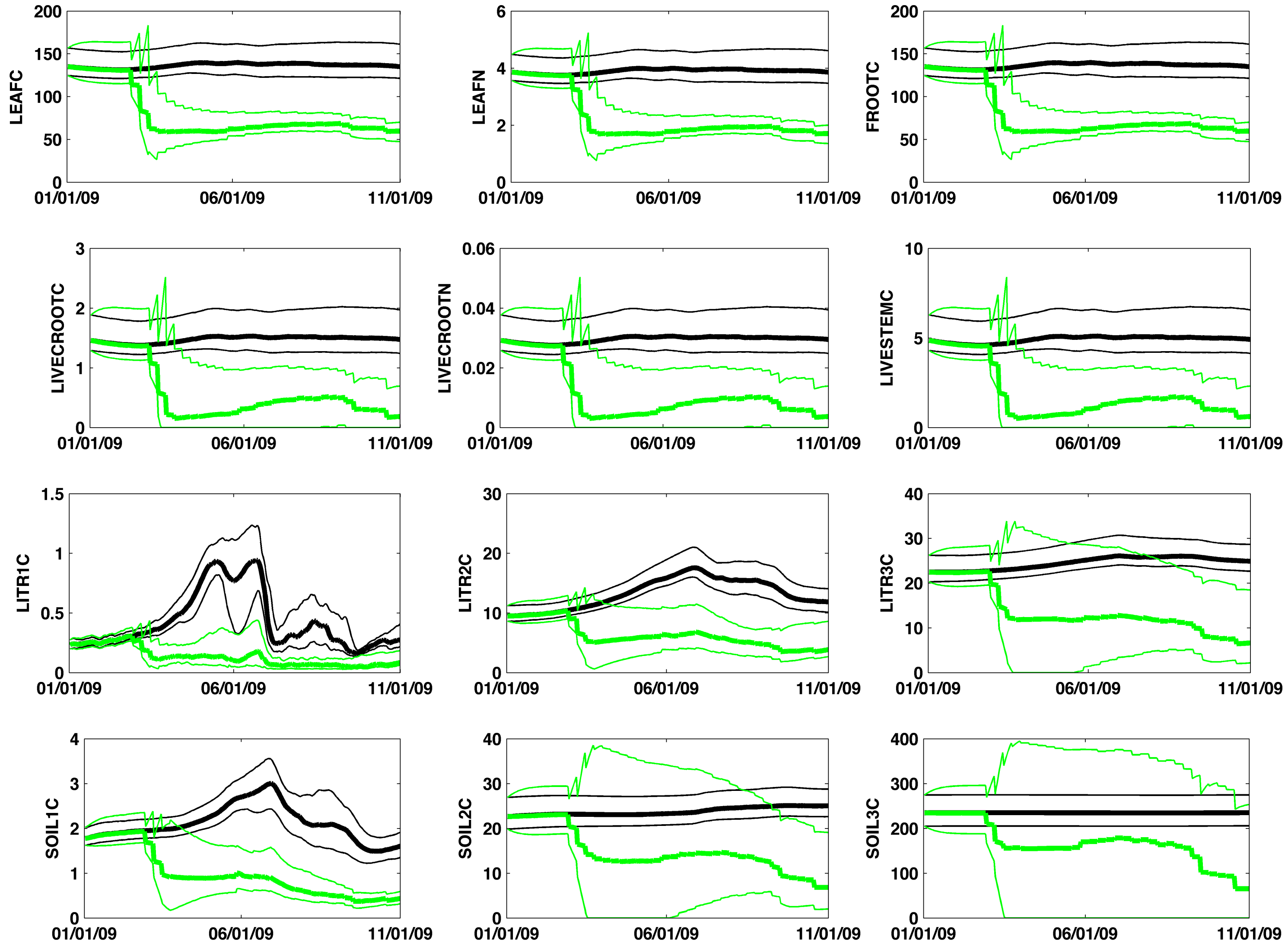


Ameriflux and MODIS LAI observations

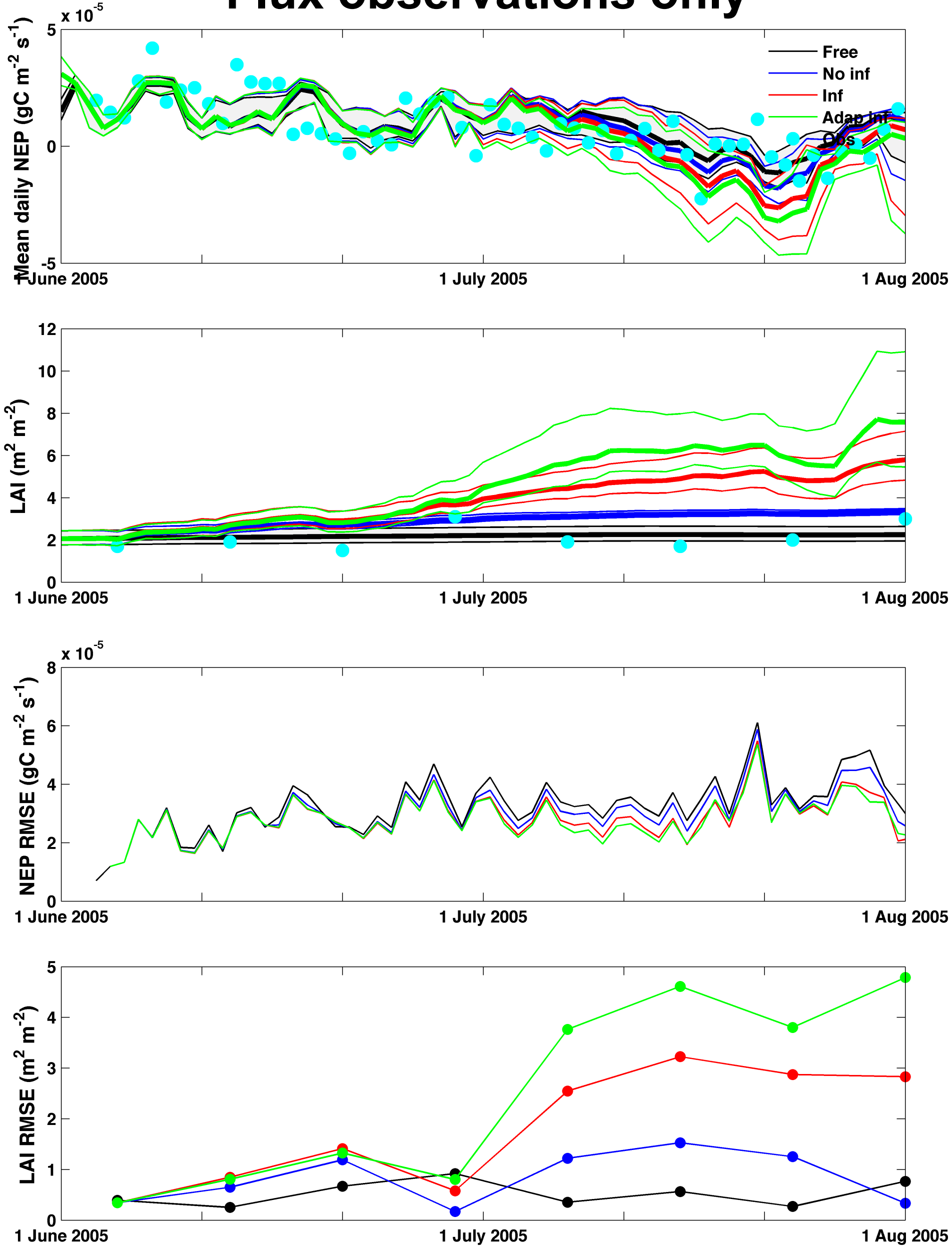




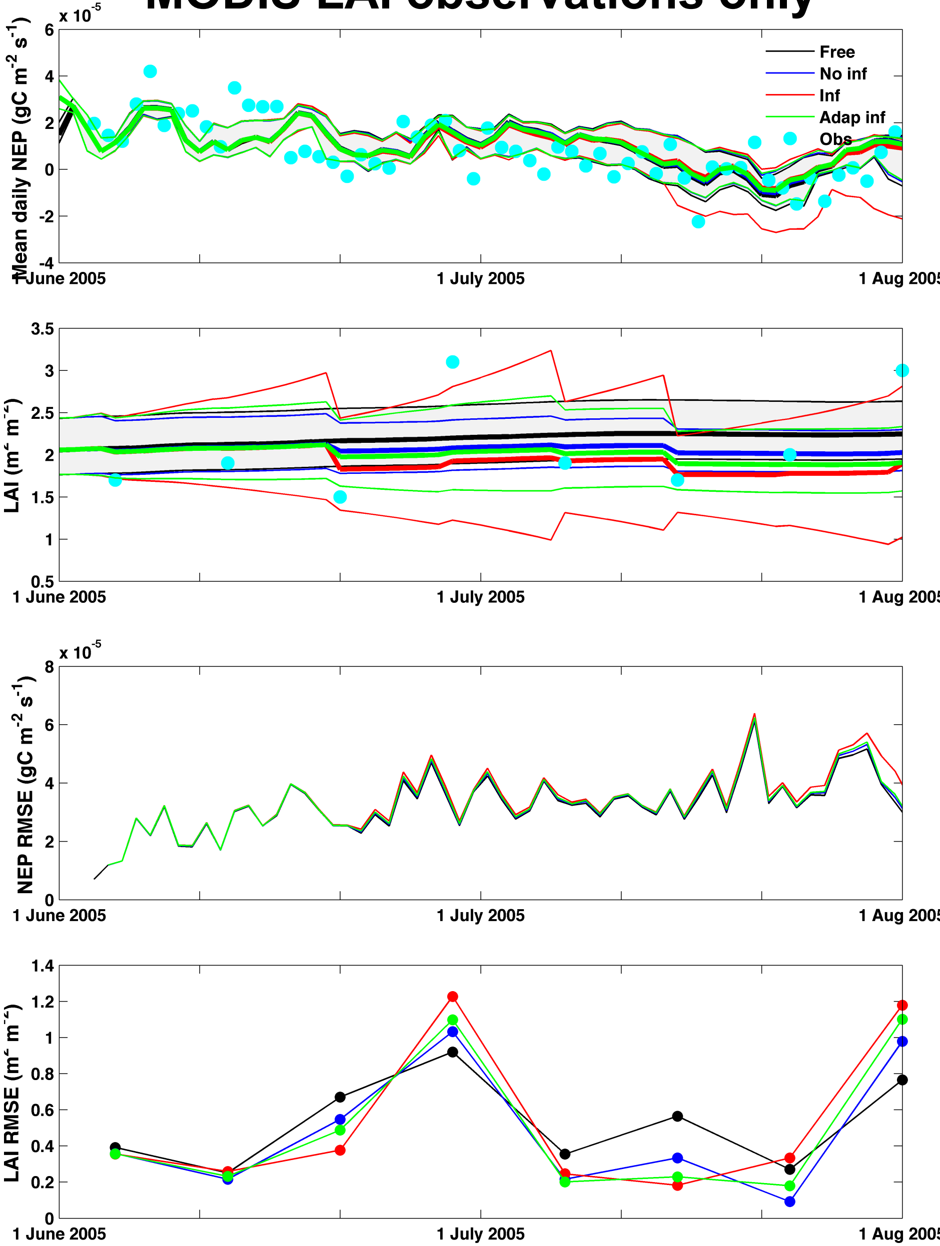
Ameriflux and MODIS LAI observations



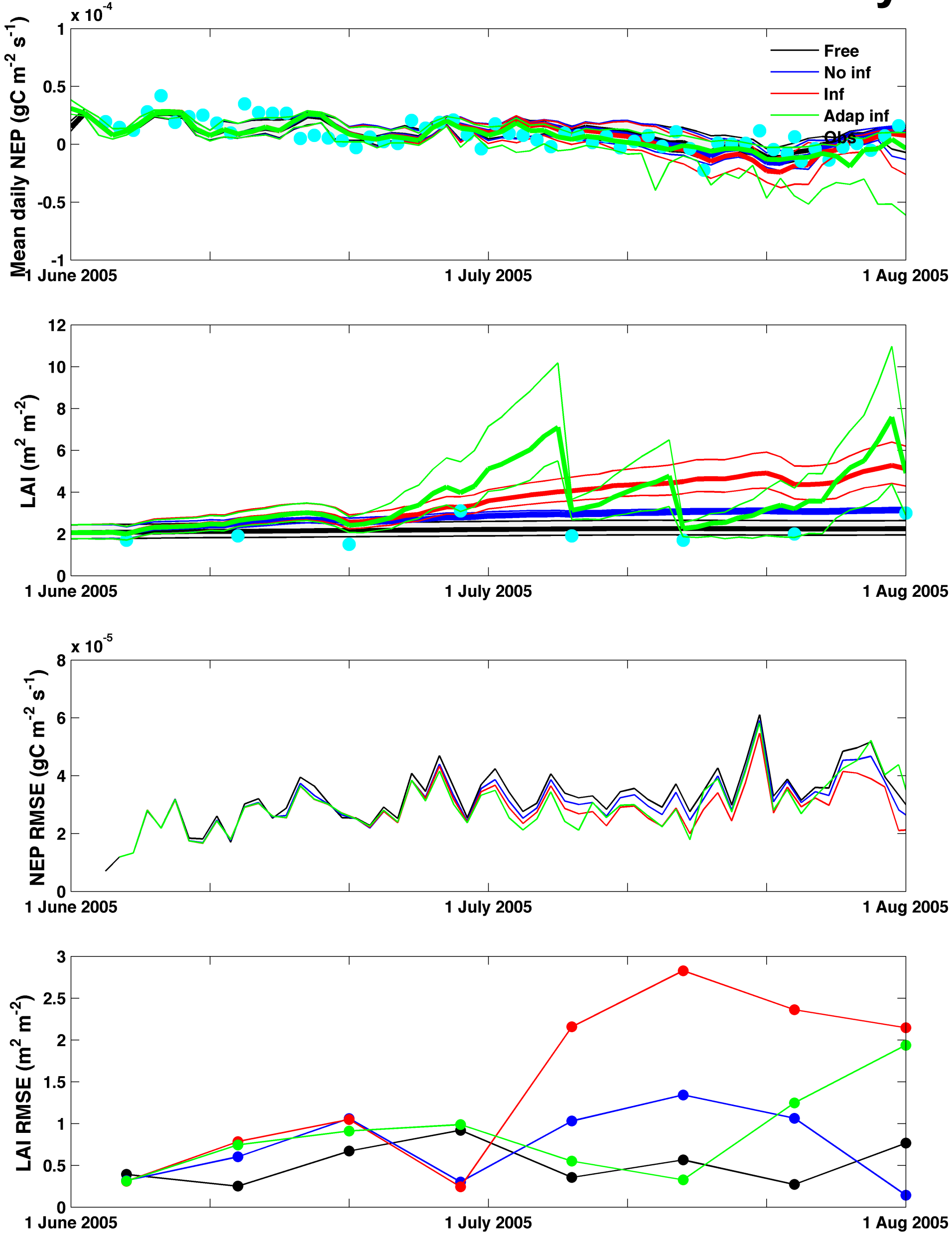
Flux observations only

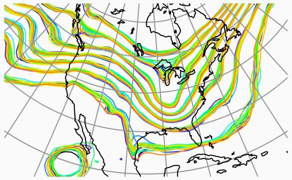


MODIS LAI observations only



Flux & MODIS LAI observations only





Open-Source Data Assimilation for Land Models and Multiscale Observations.

T. Hoar[1], A. Fox[2], Y. Zhang[3], R. Rosolem[4], A. Toure[5], B. Evans[6], J. McCreight[1]

[1]National Center for Atmospheric Research, Boulder, Colorado, USA [2]National Ecological Observatory Network, Boulder, Colorado, USA
[3]University of Texas at Austin, Austin, Texas, USA [4]University of Bristol, Bristol, England
[5]NASA GSFC, SSAI, Greenbelt, Maryland, USA [6]Macquarie University, Sydney, NSW, Australia

dart@ucar.edu

Institute for Mathematics Applied to Geosciences

1. Introduction

The Data Assimilation Research Testbed (DART) is an open source community software facility for ensemble data assimilation developed at the National Center for Atmospheric Research (NCAR). DART works with a wide variety of models and observations. Building an interface between DART and a new model does not require an adjoint and generally requires no modifications to the model code.

DART works with several land models, including:

- the Community Land Model – CLM,
- the NOAH-LSM,
- the Weather Research and Forecasting Model Hydrological modeling extension package – WRF-Hydro, and
- the Community Atmosphere-Biosphere-Land Exchange (CABLE) model.

DART assimilates dozens of observation types from a variety of sources. Some of the observations of interest for land assimilation are:

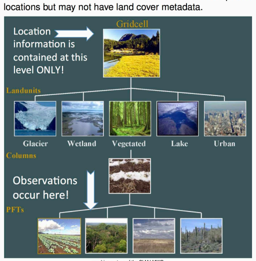
- in-situ measurements of soil moisture, temperature,
- tower fluxes,
- leaf area index,
- total water storage anomalies (i.e. GRACE),
- cosmic ray neutron intensities,
- and microwave brightness temperatures (Tb).

Anderson, J. L., et al., 2009
The Data Assimilation Research Testbed: A Community Data Assimilation Facility. *BAMS* **90** No. 9 pp. 1283–1296

http://www.image.ucar.edu/DAReS/DART has information about how to download and install DART, a full DART tutorial (included with the distribution), and how to contact us.

1.1 Land Model Structure Complications

Many land models divide gridcells into proportional units based on land cover characteristics. This is a challenge for data assimilation as the land units generally have no unique location information of their own. Observations have specific locations but may not have land cover metadata.



1.2 Observation Metadata

All ensemble data assimilation systems require the ability to calculate the expected value of the observation given a model state. The accurate application of this calculation (the observation operator) may require:

- knowledge of what land cover unit(s) or PFT to use for the calculation,
- soil properties, and
- instrument-specific parameters

Some of these could come from a lookup table based on PFT or location, but the lookup table generally must be recomputed to match the model resolution. Some (like Tb polarizations and frequencies) must be part of each observation.

Fall AGU 2014

2. Ecological State Estimation

Andrew Fox *National Ecological Observatory Network*

In an observation system simulation experiment (OSSE) we treat one ensemble member as "truth" and sample with appropriate noise at 60 NEON site locations to observe Leaf Area Index (LAI) every 8 days, Leaf Nitrogen every 12 days, and Net Ecosystem Productivity and Evapotranspiration every 0.5 hours. We then investigate the impacts of assimilating ≈ 520,000 synthetic observations over a 3 month period.

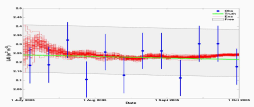


Figure 1: *This "sawtooth" plot shows LAI simulated by all 80 ensemble members in a grid cell with observations. The increments (updates) calculated by the filter move the ensemble towards the observations and result in a reduction in uncertainty (spread) around the truth. In this case, uncertainty is reduced too much and the result is slightly biased.*

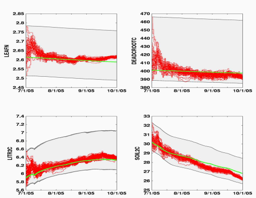


Figure 2: *The DART state vector contains more than 20 variables, including all the large carbon and nitrogen pools. These can all be updated by the filter through their covariance with observed variables. The allocation algorithms in CLM mean observations provide a strong constraint on many unobserved variables.*

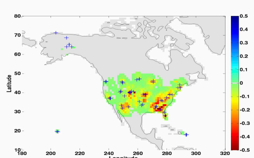


Figure 3: *Change in LAI spread in posterior ensemble, 31 July 2005. The largest innovations are near the observations, but not necessarily in the exact grid cell. Carbon pools from all grid cells are in the DART state vector and information can propagate from sites to regions. A cutoff value limits the distance over which this can occur.*

3. Multisensor Assimilation

Yongfei Zhang, *University of Texas at Austin.*

The DART algorithms can assimilate observations with uncorrelated observation errors in any order (Anderson, 2003). This allows one to simultaneously assimilate MODIS/Terra snow cover fraction (SCF – with little information about snow amount) and Gravity Recovery and Climate Experiment (GRACE) estimates of total water storage anomalies.

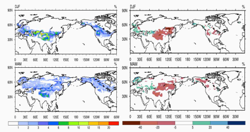


Figure 4: *Left: Ensemble spread of SCF for (top) DJF and (bottom) MAM in 2002-2003. Ensemble spread is calculated as the standard deviation of SCF among 40 ensemble members. Right: The difference of SCF between the data assimilation case and the open loop case averaged for (top) DJF and (bottom) MAM.*

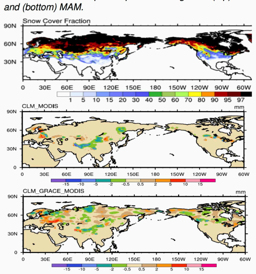


Figure 5: *Top: Mean SCF for Dec 2002 through Feb 2003. Middle: Impact (assimilation minus open loop) on snow water equivalent for 15 March 2003 for MODIS-only assimilation. Bottom: Impact on SCF for an assimilation with both MODIS and GRACE observations. GRACE is clearly providing additional information in regions where the SCF is saturated.*

Zhang, Y.-F., et al., 2014 *Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4* DOI: 10.1002/2013JD021329

4. Brightness Temperature Observations

Ally M. Toure, *NASA GSFC, USRA.*

The objective is to assess the performance of the brightness temperature (Tb) prediction in the Community Land surface Model version 4 (CLM4) coupled with a snow Radiative Transfer Model (RTM).

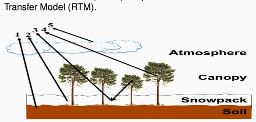


Figure 6: *The main contributions to the microwave emission measured by a spaceborne radiometer: 1) upward emitted soil emission, 2) snowpack, 3) combined canopy and snowpack, 4) canopy, and 5) the atmosphere.*

The RTM used to predict Tb using the CLM4 output is the Microwave Emission Model of Layered Snowpacks (MEMLS). It simulates Tbs for multi-layer snowpack and is valid for the frequency range of 5 GHz to 100 GHz. Typical inputs to the model are: Tb : the snowpack brightness temperature, S_0 : the ground-snow interface reflectivity, T_0 : the ground temperature, S_j : the interface reflectivity on top of each snow layer j, d_j : layer thickness, T_j : layer temperature, r_j : layer internal reflectivity, e_j : layer emissivity, t_j : transmissivity of each layer, and T_{sky} : the downwelling (sky) radiation.

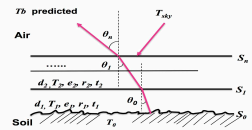


Figure 7: *Schematic of the MEMLS snow RTM (Wiesmann and Mätzler, 1999).*

RTM's also have parameters that must be estimated and are usally spatially varying. (See, for example, De Lannoy et. al, 2013: Global Calibration of the GEOS-5 L-Band Microwave Radiative Transfer Model over Nonfrozen Land Using SMOS Observations. J. Hydrometeor, 14, 765-785. doi: http://dx.doi.org/10.1175/JHM-D-12-092.1)

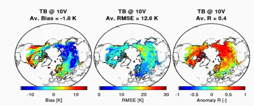


Figure 8: *Annual mean bias, RMSE and anomaly correlation coefficient (R) between predicted Tb and AMSR-E observations during 2002-2010. This is just 1 of 6 frequencies, and 1 polarization (10.7 Ghz, V pol).*

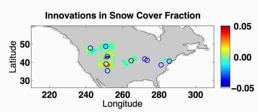


Figure 9: *This figure shows the amount of change induced in the snow cover fraction from assimilating synthetic brightness temperatures at 10 locations (shown as ◯) for 31 January 2001.*

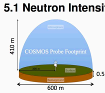
Tb assimilation presents other problems for DA. If you were to use all 6 AMSR-E frequencies at both polarizations for both ascending and descending swaths, you would be assimilating more than 6,000,000 observations *per day* in the Northern Hemisphere alone!

5. Soil Moisture Observations

Rafael Rosolem, *University of Bristol.*

DART has been coupled to the NOAH Land Surface Model (HRLDAS-V3.3) and provides an operator to return neutron intensity "observations" given a soil moisture profile. This can be used to update the NOAH model state.

5.1 Neutron Intensity Observations



The COSMOS probe measures the neutron intensity for a given volume, which is related to the amount of Hydrogen present. The COSMIC model relates the neutron intensity to total soil moisture.

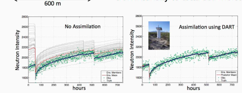


Figure 10: *An early result for the Santa Rita site.*

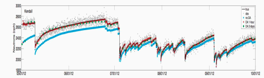


Figure 11: *Expected neutron intensity (counts per hour – cph) from a set of 3 experiments. The (known) true state is a solid black line, realistic observations (i.e. truth plus noise) are indicated by the gray dots, blue diamonds are a free run (i.e. no DA), red dots are assimilation every hour, green squares are assimilation every 2 days.*

5.2 Real Observations

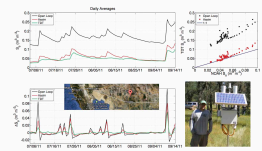


Figure 12: *These graphics assess the performance of the assimilation of neutron intensity observations on soil moisture to withheld traditional soil moisture observations. The posterior mean is plotted in red.*

Rosolem, R., et al., 2014 *Translating above-ground cosmic-ray neutron intensity to high-frequency soil moisture profiles at sub-kilometer scale.* *HESS* **18**, pp. 4363-4379

6. Hydrologic Assimilation

James McCreight, *NCAR.*

The Weather Research and Forecasting Model Hydrological modeling extension package (WRF-Hydro) is a community-based model coupling framework designed to link multi-scale process models of the atmosphere and terrestrial hydrology. Research with DART and WRF-Hydro will enable: 1) improved forecasts by reducing error in initial conditions, 2) a high-quality reanalysis, 3) diagnosis of model or observation errors, and 4) exploration of targeted observations.

6.1 Streamflow Assimilation

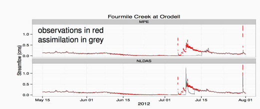


Figure 13: *Simulation before including several parameters into the DA for the two sets of precipitation forcing data. Top: precipitation from NOAA's Multisensor Precipitation Estimate (MPE), Bottom: NLDAS precipitation. The assimilation result runoff (grey) is highly dependent on the forcing.*

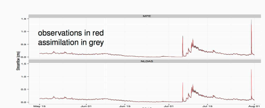


Figure 14: *Simulation after including several parameters into the DA for the two sets of forcing data.*

7. Land Surface Data Assimilation – CABLE

Brad Evans, *TERN/eMAST* **Luigi Renzullo**, *CSIRO*

Researchers from CSIRO, Macquarie University and the National Computing Infrastructure (NCI) teamed up with US collaborators to install and run DART on NCI's supercomputer (Raijin) and coupled it to Australia's Community Atmosphere Biosphere Land Exchange (CABLE) land surface model. The endeavour marks significant progress toward the vision of the Ecosystem Modelling and Scaling Infrastructure (eMAST) facility under the Terrestrial Ecosystem Research Network (TERN) to develop Australia's first modelling and data integration system for ecosystem science and monitoring at unparalleled scales in space and time. The system brings together a range of disparate ecological observations from ground- and space-based sensing networks into CABLE's modelling framework.

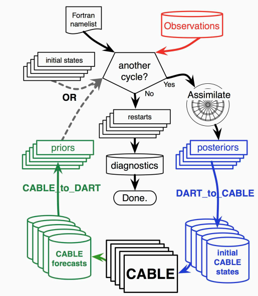


Figure 15: *A schematic of the assimilation system with DART and CABLE. Starting at the top: DART reads in an initial ensemble, the observations, the run-time control information and performs an assimilation to create posterior estimates of the CABLE variables.* DART_TO_CABLE *conveys the posteriors to a set of CABLE restart files which are advanced by CABLE to the time of the next observation.* CABLE_TO_DART *then extracts the prognostic state variables of interest and converts them to a DART-compliant format.*

Some of the instruments providing the observations that can be assimilated in the CABLE/DART system. Left-to-right: Eddy Covariance (Cape Tribulation), OzFlux (Scott Farm), CosmOz (Tullochgorum).

NCRIS National Research Infrastructure for Australia. An Australian Government Initiative. The Terrestrial Ecosystem Research Network Ecosystem Modelling and Scaling Infrastructure is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

Anderson, J.-L., 2003
A local least squares framework for ensemble filtering. *MWR* **131**, pp. 634-642

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2013JD021329

Key Points:
- This work interfaced CLM4 with DART
- MODIS snow cover is assimilated into DART/CLM4
- The RMSE of snow cover and snow depth is reduced

Correspondence to:
Z.-L. Yang,
liang@jsg.utexas.edu

Citation:
Zhang, Y.-F., T. J. Hoar, Z.-L. Yang, J. L. Anderson, A. M. Toure, and M. Rodell (2014), Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4, *J. Geophys. Res. Atmos.*, *119*, 7091–7103, doi:10.1002/2013JD021329.

Received 7 DEC 2013
Accepted 28 MAY 2014
Accepted article online 2 JUN 2014
Published online 18 JUN 2014

Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4

Yong-Fei Zhang[1], Tim J. Hoar[2], Zong-Liang Yang[1], Jeffrey L. Anderson[2], Ally M. Toure[3,4], and Matthew Rodell[4]

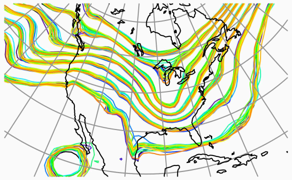
[1]Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA, [2]National Center for Atmospheric Research, Boulder, Colorado, USA, [3]Universities Space Research Association, Columbia, Maryland, USA, [4]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Abstract To improve snowpack estimates in Community Land Model version 4 (CLM4), the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) was assimilated into the Community Land Model version 4 (CLM4) via the Data Assimilation Research Testbed (DART). The interface between CLM4 and DART is a flexible, extensible approach to land surface data assimilation. This data assimilation system has a large ensemble (80-member) atmospheric forcing that facilitates ensemble-based land data assimilation. We use 40 randomly chosen forcing members to drive 40 CLM members as a compromise between computational cost and the data assimilation performance. The localization distance, a parameter in DART, was tuned to optimize the data assimilation performance at the global scale. Snow water equivalent (SWE) and snow depth are adjusted via the ensemble adjustment Kalman filter, particularly in regions with large SCF variability. The root-mean-square error of the forecast SCF against MODIS SCF is largely reduced. In DJF (December-January-February), the discrepancy between MODIS and CLM4 is broadly ameliorated in the lower-middle latitudes (23°–45°N). Only minimal modifications are made in the higher-middle (45°–66°N) and high latitudes, part of which is due to the agreement between model and observation when snow cover is nearly 100%. In some regions it also reveals that CLM4-modeled snow cover lacks heterogeneous features compared to MODIS. In MAM (March-April-May), adjustments to snow move poleward mainly due to the northward movement of the snowline (i.e., where largest SCF uncertainty is and SCF assimilation has the greatest impact). The effectiveness of data assimilation also varies with vegetation types, with mixed performance over forest regions and consistently good performance over grass, which can partly be explained by the linearity of the relationship between SCF and SWE in the model ensembles. The updated snow depth was compared to the Canadian Meteorological Center (CMC) data. Differences between CMC and CLM4 are generally reduced in densely monitored regions.

1. Introduction

Snow plays a unique role in the global hydrological cycle, water resources management, and atmospheric predictability. Its special physical properties (high albedo, low thermal conductivity, and ability to change phase) significantly modulate energy and water exchanges between the atmosphere and the land surface [*Goodison et al.*, 1999]. In regions where streamflow is dominated by snowmelt, the performance of hydrological forecasts largely depends on snowpack estimates at the beginning of the forecast period [*Clark and Hay*, 2004]. Snowpack acts as a key boundary condition for the atmosphere and influences atmospheric predictability. A more realistic simulated snowpack enhances springtime surface air temperature predictability [e.g., *Peings et al.*, 2010]. Furthermore, snowpack impacts atmospheric circulations through teleconnections. Numerous modeling and observational studies have shown an inverse relationship between the winter and springtime Eurasian snow-covered area and the summertime Indian Monsoon rainfall [e.g., *Vernek et al.*, 1995; *Bamzai and Shukla*, 1999; *Turner and Slingo*, 2011].

A variety of snowpack products have been generated for hydroclimatic analysis and evaluation of climate models. Ground measurements usually lack spatial representativeness, especially in regions of high heterogeneity [*Liston*, 2004], and are difficult to obtain in many regions especially in complex terrains; therefore, satellite remote sensing plays an important role in producing global snowpack estimates. Based on the optical properties of snow, observations of visible and near-infrared bands can detect snow extent in



1. Introduction

The Data Assimilation Research Testbed (DART) is an open source community software facility for ensemble data assimilation developed at the National Center for Atmospheric Research (NCAR). DART works with a wide variety of models and observations. Building an interface between DART and a new model does not require an adjoint and generally requires no modifications to the model code.

DART works with several land models, including:
- the Community Land Model – CLM,
- the NOAH-LSM,
- the Weather Research and Forecasting Model Hydrological modeling extension package – WRF-Hydro, and
- the Community Atmosphere-Biosphere-Land Exchange (CABLE) model.

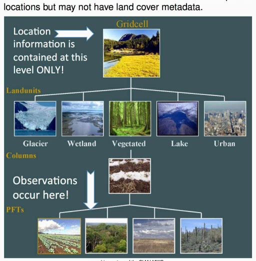
DART assimilates dozens of observation types from a variety of sources. Some of the observations that are used for land assimilation are:
- in-situ measurements of soil moisture, temperature,
- tower fluxes,
- leaf area index,
- total water storage anomalies (i.e. GRACE),
- cosmic ray neutron intensities,
- and microwave brightness temperatures (Tb).

Anderson, J. L., et al., 2009 *The Data Assimilation Research Testbed: A Community Data Assimilation Facility.* *BAMS* **90** No. 9 pp. 1283–1296

http://www.image.ucar.edu/DAReS/DART has information about how to download and install DART, a full DART tutorial (included with the distribution), and how to contact us.

1.1 Land Model Structure Complications

Many land models divide gridcells into proportional units based on land cover characteristics. This is a challenge for data assimilation as the land units generally have no unique location information of their own. Observations have specific locations but may not have land cover metadata.



graphic courtesy of the CLM LMWG

1.2 Observation Metadata

All ensemble data assimilation systems require the ability to calculate the expected value of the observation given a model state. The accurate application of this calculation (the observation operator) may require:
- knowledge of what land cover unit(s) or PFT to use for the calculation,
- soil properties, and
- instrument-specific parameters

Some of these could come from a lookup table based on PFT or location, but the lookup table generally must be recomputed to match the model resolution. Some (like Tb polarizations and frequencies) must be part of each observation.

Fall AGU 2014

7. Land Surface Data Assimilation – CABLE

Brad Evans, *TERN/eMAST* **Luigi Renzullo**, *CSIRO*

Researchers from CSIRO, Macquarie University and the National Computing Infrastructure (NCI) teamed up with US collaborators to install and run DART on NCI's supercomputer (Raijin) and coupled it to Australia's Community Atmosphere Biosphere Land Exchange (CABLE) land surface model. The endeavour marks significant progress toward the vision of the Ecosystem Modelling and Scaling Infrastructure (eMAST) facility under the Terrestrial Ecosystem Research Network (TERN) to develop Australia's first modelling and data integration system for ecosystem science and monitoring at unparalleled scales in space and time. The system brings together a range of disparate ecological observations from ground- and space-based sensing networks into CABLE's modelling framework.

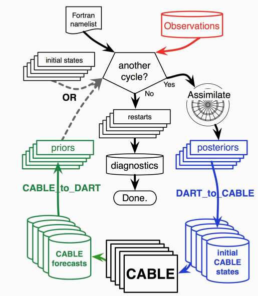


Figure 15: *A schematic of the assimilation system with DART and CABLE. Starting at the top: DART reads in an initial ensemble, the observations, the run-time control information and performs an assimilation to create posterior estimates of the CABLE variables.* DART TO CABLE *conveys the posteriors to a set of CABLE restart files which are advanced by CABLE to the time of the next observation.* CABLE TO DART *then extracts the prognostic state variables of interest and converts them to a DART-compliant format.*

Some of the instruments providing the observations that can be assimilated in the CABLE/DART system. Left-to-right: Eddy Covariance (Cape Tribulation), OzFlux (Scott Farm), CosmOz (Tullochgorum).

The Terrestrial Ecosystem Research Network Ecosystem Modelling and Scaling Infrastructure is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

Anderson, J.-L., 2003 *A local least squares framework for ensemble filtering.* *MWR* **131**, pp. 634-642

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

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?
- What are the best temporal aggregation strategies for EC flux tower data?
- Can we develop appropriate observation 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?

Future Directions

- Optimizing NEON data delivery for use with land models

Constant interaction with the modeling community

- NEON will provide systematic observations sampling a wide climate space to constrain models in a variety of ways

Community model development and improvement

- Community tools for data assimilation provides a means of directly utilizing this new information

Community development of DA techniques with land models leading to improvements in forecasts

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