CTSM and the Data Assimilation Research Testbed

Andrew Fox[1,2], and Tim Hoar[2], and the DART Team

1. University of Arizona 2. National Center for Atmospheric Research

1. WHY DO WE WANT TO DO THIS?

2. WHAT WE CAN DO…

3. HOW WE ACTUALLY DO THIS…

4. WHAT ARE WE THINKING ABOUT NEXT?

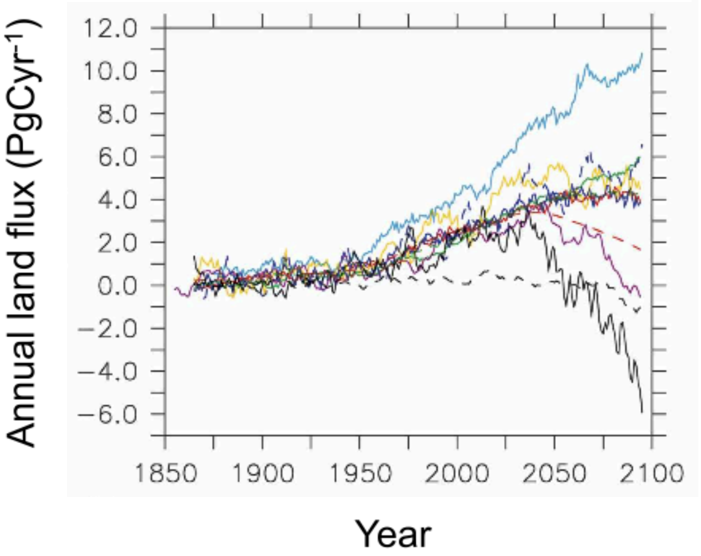
Huge uncertainties in future land carbon flux

VOLUME 19 JOURNAL OF CLIMATE 15 JULY 2006

2006

Climate–Carbon Cycle Feedback Analysis: Results from the C^4MIP 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]



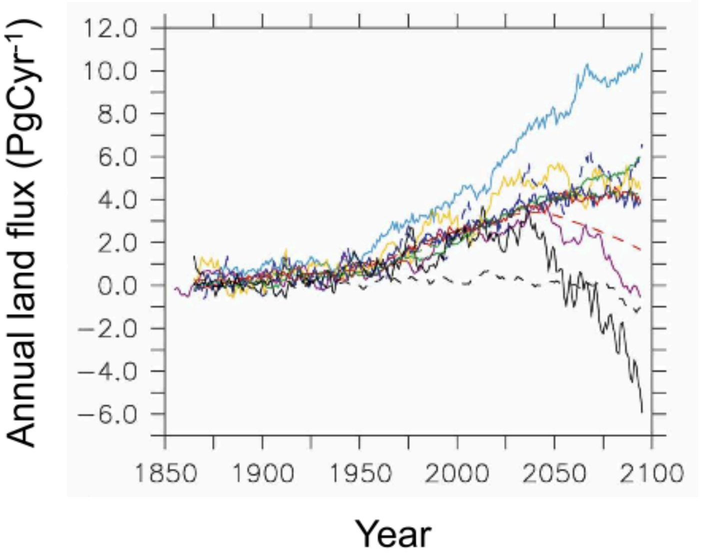
Future uncertainties have not been reduced

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]



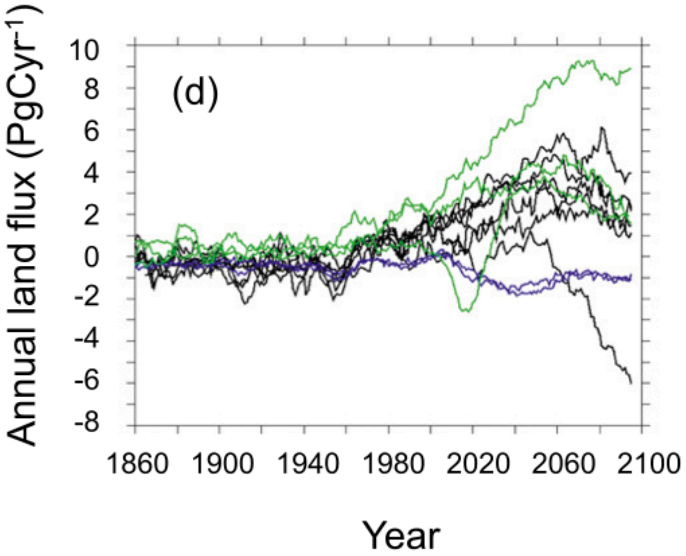
15 January 2014 FRIEDLINGSTEIN ET AL. 511

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[&]



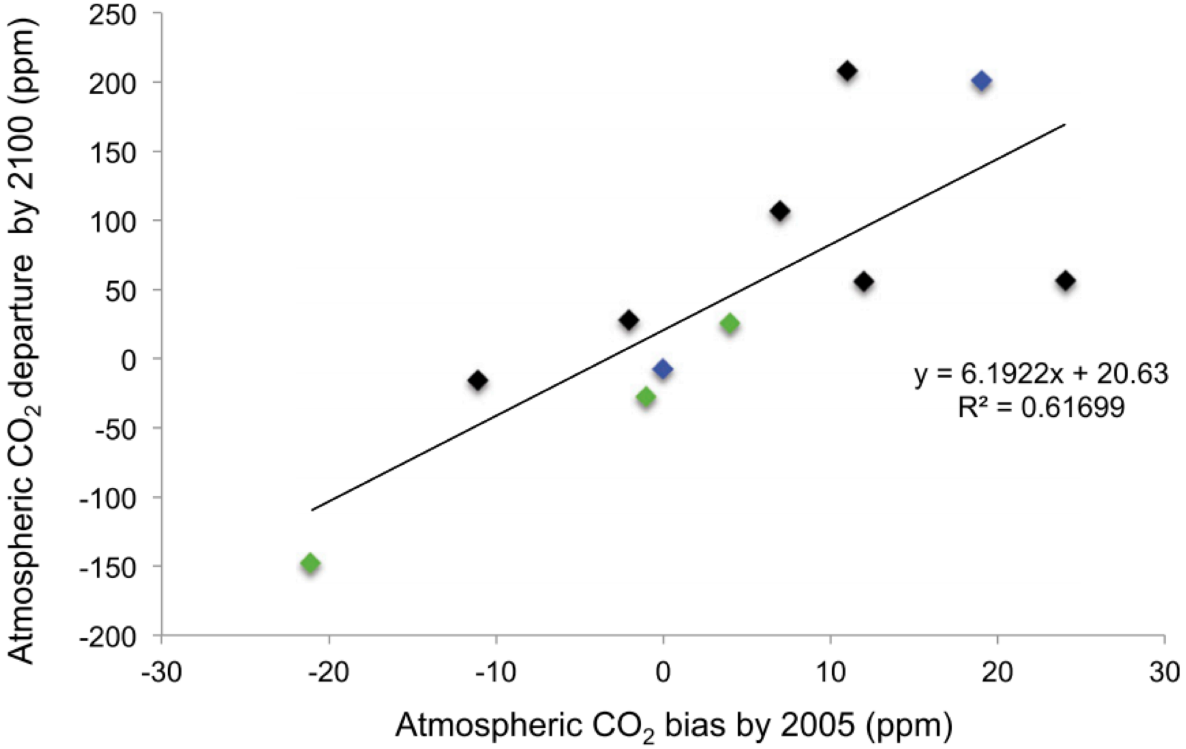
NCAR The University of Arizona

Sources of uncertainty (& their cure)

- Model Structure — DEVELOPMENT
- Model Parameters — PARAMETERIZATION
- Initial Conditions
- Spin Up
- Climate forcing

} (STATE) DATA ASSIMILATION

Constrain the future by constraining the present



Friedlingstein et al., 2014

FIG. 3. Relationship between model bias in simulating present-day (2005) atmospheric CO_2, and the difference between 2100 simulated CO_2 and baseline estimate from MAGICC6

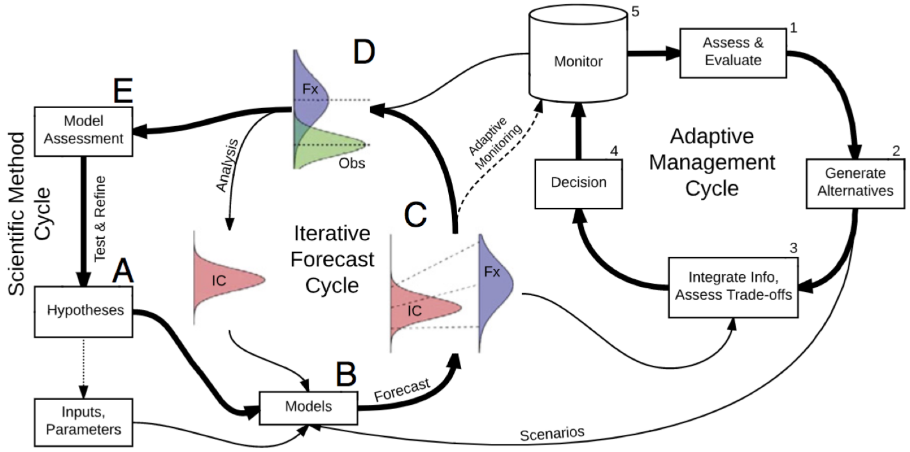
NCAR THE UNIVERSITY OF ARIZONA

Particularly important over shorter timescales

PNAS

Iterative near-term ecological forecasting: Needs, opportunities, and challenges

Michael C. Dietze[a,1], Andrew Fox[b], Lindsay M. Beck-Johnson[c], Julio L. Betancourt[d], Mevin B. Hooten[e,f,g], Catherine S. Jarnevich[h], Timothy H. Keitt[i], Melissa A. Kenney[j], Christine M. Laney[k], Laurel G. Larsen[l], Henry W. Loescher[k,m], Claire K. Lunch[k], Bryan C. Pijanowski[n], James T. Randerson[o], Emily K. Read[p], Andrew T. Tredennick[q,r], Rodrigo Vargas[s], Kathleen C. Weathers[t], and Ethan P. White[u,v,w]

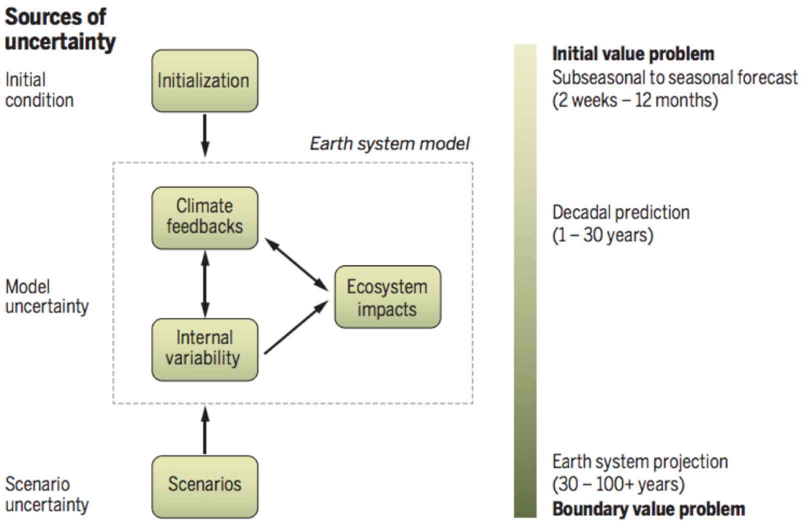


REVIEW

EARTH SYSTEMS

Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models

Gordon B. Bonan[1]* and Scott C. Doney[2]*



NCAR THE UNIVERSITY OF ARIZONA

What is Data Assimilation?

DATA
+ Often high quality, relevant data
+ Clear uncertainties
- Limited spatial extent
- Limited temporal span

What is Data Assimilation?

DATA
+ Often high quality, relevant data
+ Clear uncertainties
- Limited spatial extent
- Limited temporal span

MODELS
+ Extrapolation and forecasts
+ Provide system understanding
- Subjective simplifications
- Uncertainties difficult to quantify

What is Data Assimilation?

DATA
+ Often high quality, relevant data
+ Clear uncertainties
- Limited spatial extent
- Limited temporal span

MODELS
+ Extrapolation and forecasts
+ Provide system understanding
- Subjective simplifications
- Uncertainties difficult to quantify

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

What is Data Assimilation?

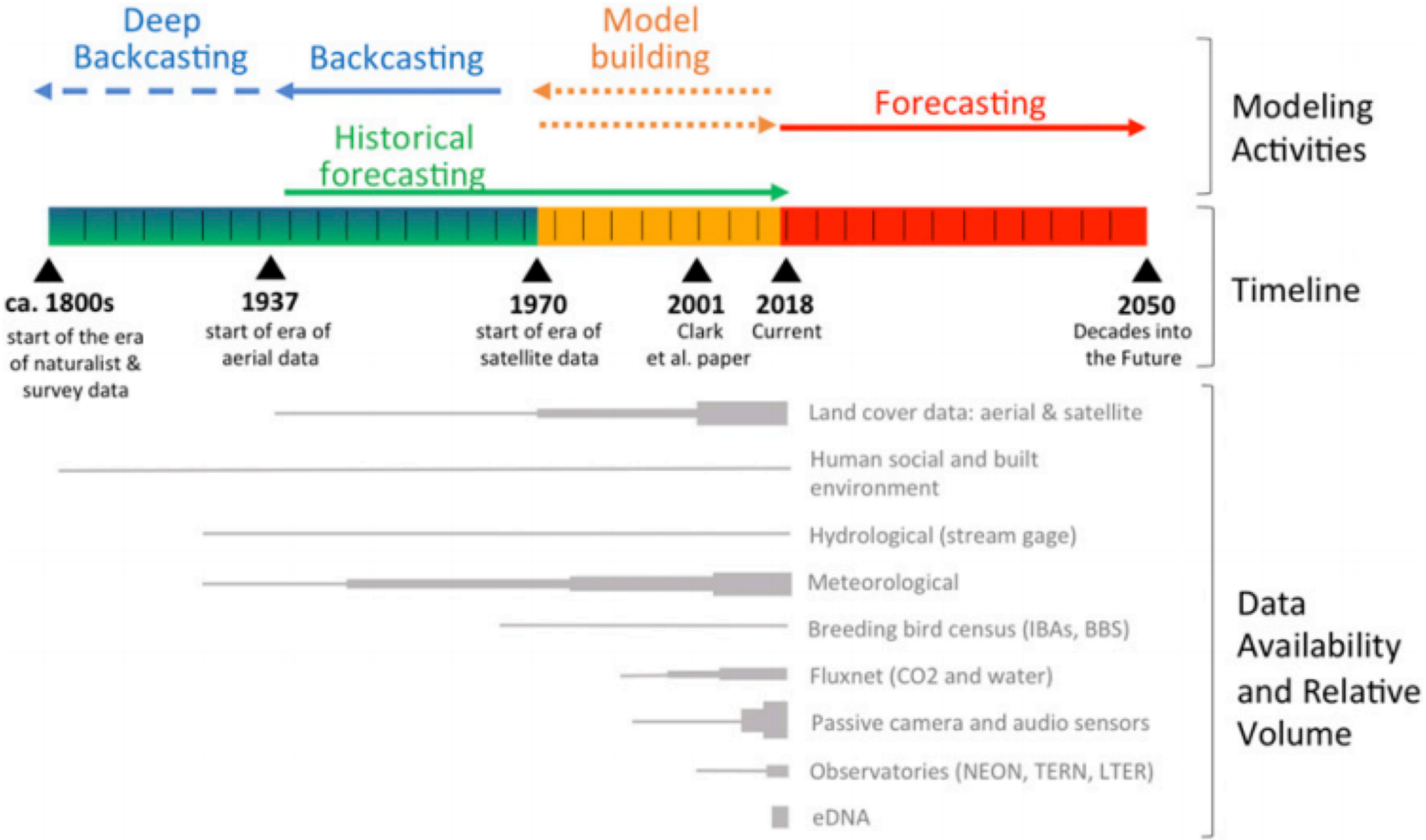
DATA
+ Often high quality, relevant data
+ Clear uncertainties
- Limited spatial extent
- Limited temporal span

MODELS
+ Extrapolation and forecasts
+ Provide system understanding
- Subjective simplifications
- Uncertainties difficult to quantify

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

ANALYSIS
+ Spatially complete
+ Consistent with observations
+ Quantified uncertainties
+ More accurate forecasts

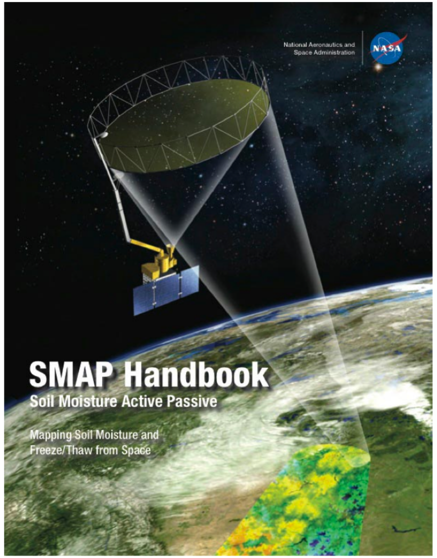
Data availability increasing rapidly

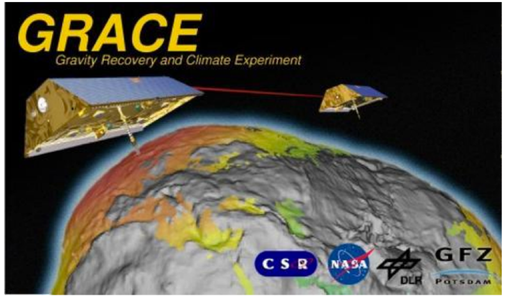


PNAS 2018; published ahead of print January 30, 2018, https://doi.org/10.1073/pnas.1710231115

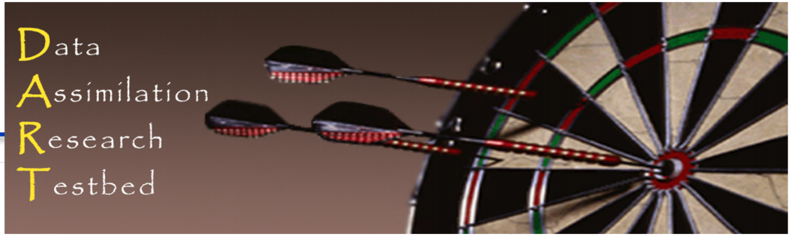
NCAR THE UNIVERSITY OF ARIZONA

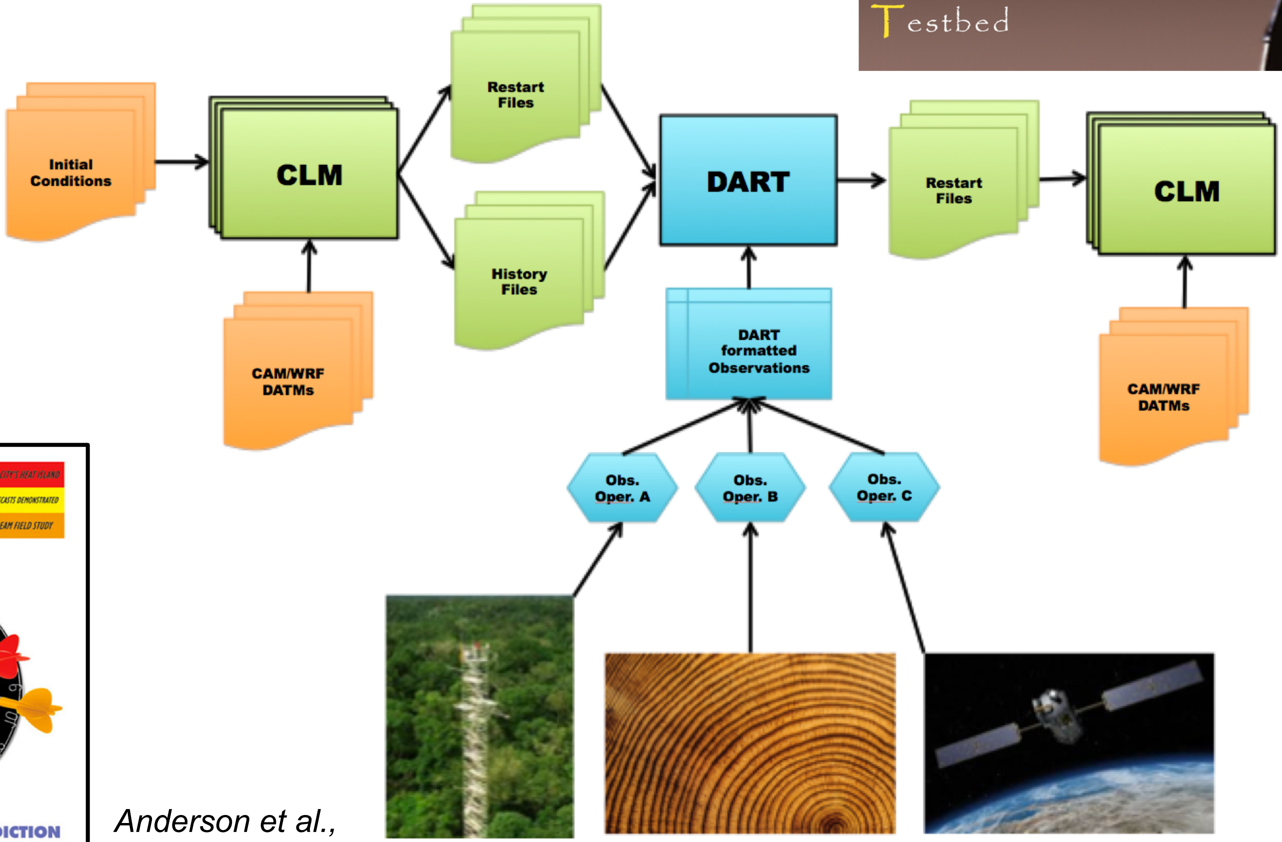
Remote Sensing Observations are key

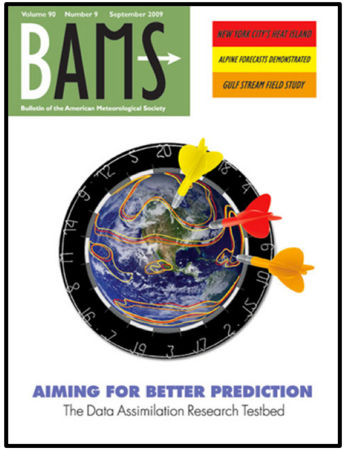




CLM-DART







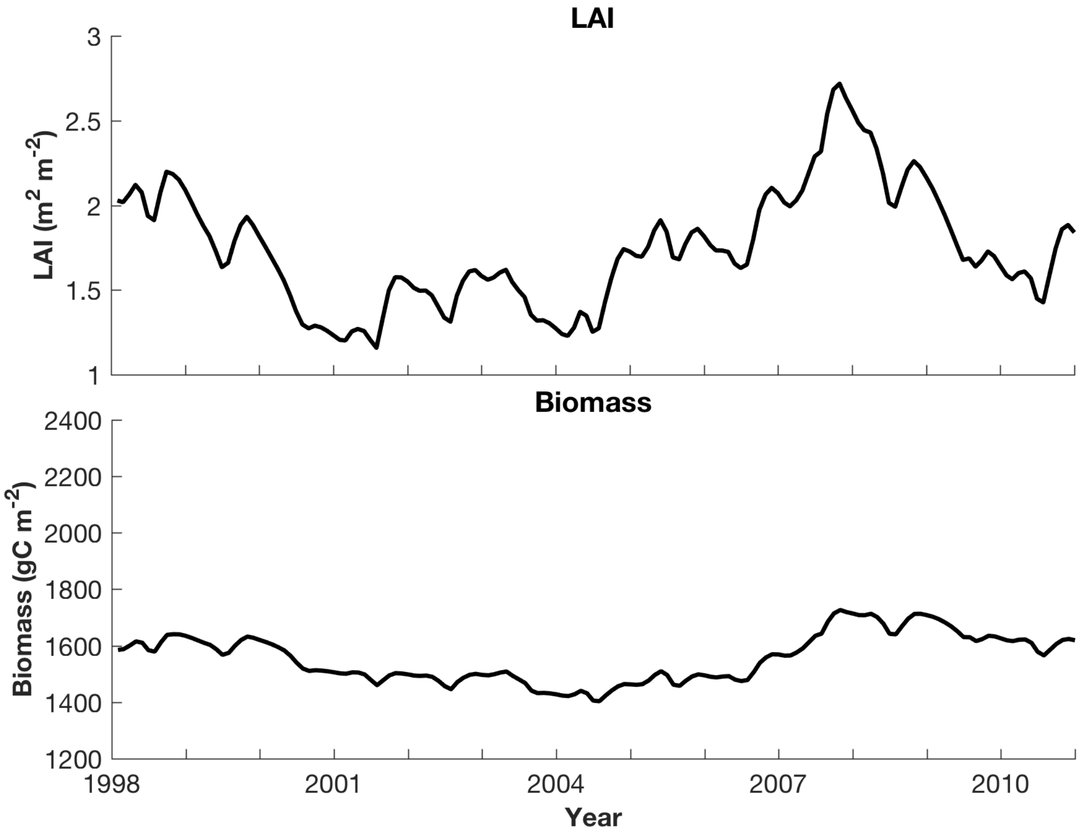
Anderson et al., 2009

NCAR THE UNIVERSITY OF ARIZONA

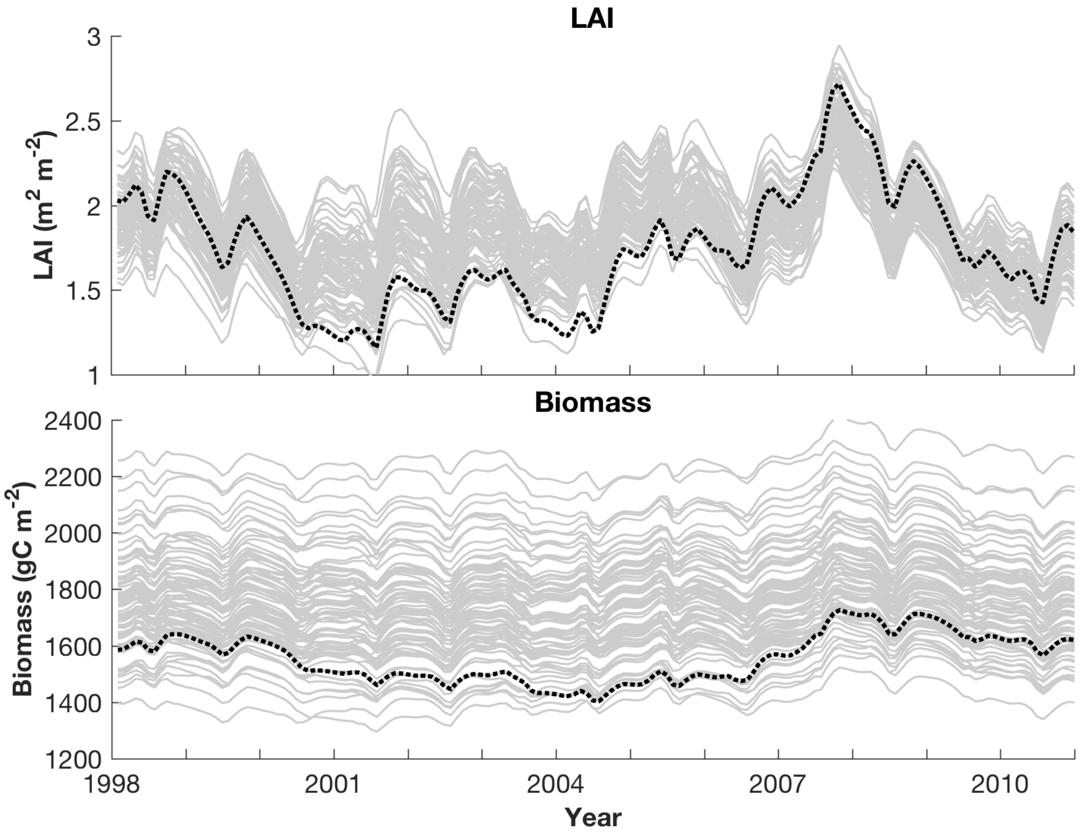
WHAT WE CAN DO…

Examples using CTSM-DART

LAI and Biomass – single instance



LAI and Biomass – multi-instance



LAI and Biomass – Observations

Bi-weekly, 0.5° Aggregated MODIS LAI Observations

Remote Sens. **2013**, *5*, 927-948; doi:10.3390/rs5020927

OPEN ACCESS
Remote Sensing
ISSN 2072-4292
www.mdpi.com/journal/remotesensing

Article

Global Data Sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g Derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the Period 1981 to 2011

Zaichun Zhu [1,2,*,†], Jian Bi [1,†], Yaozhong Pan [2], Sangram Ganguly [3], Alessandro Anav [4], Liang Xu [1], Arindam Samanta [5], Shilong Piao [6,7], Ramakrishna R. Nemani [8] and Ranga B. Myneni [1]

Annual, 0.25° Vegetation Optical Depth Biomass Observations

nature climate change

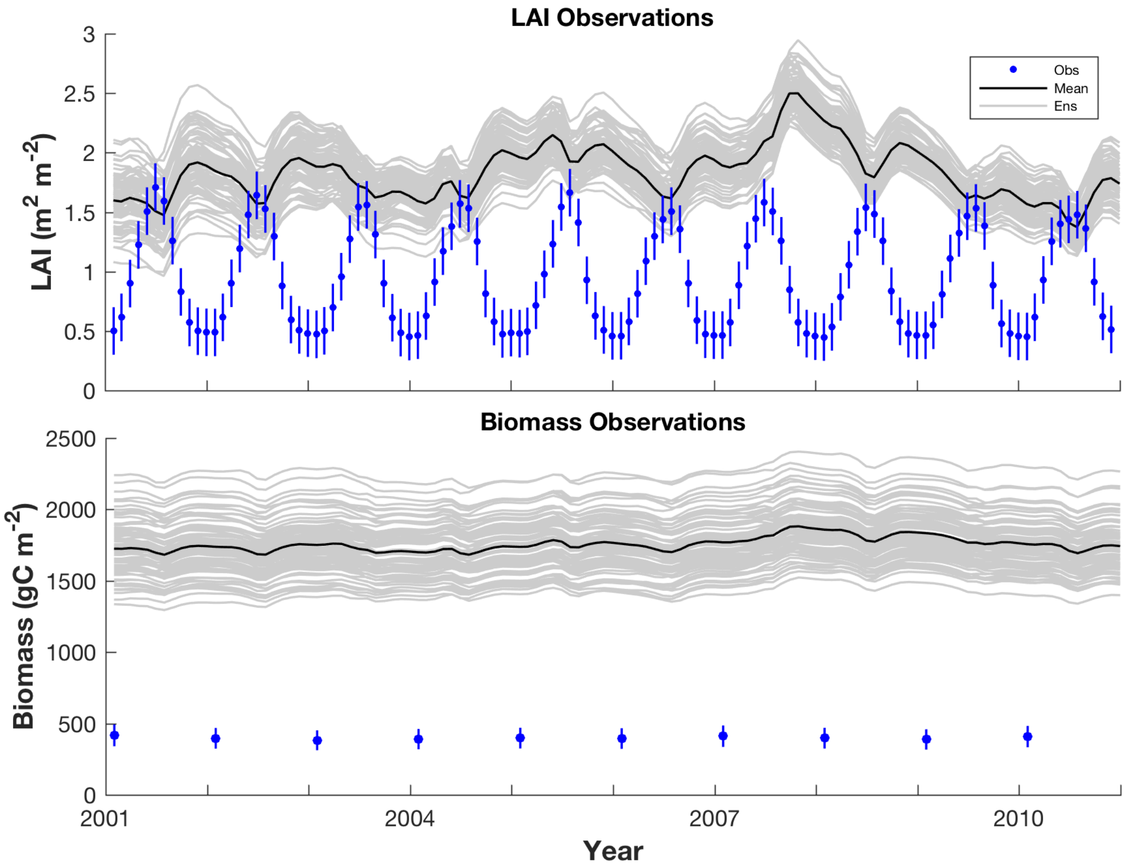
LETTERS

PUBLISHED ONLINE: 30 MARCH 2015 | DOI: 10.1038/NCLIMATE2581

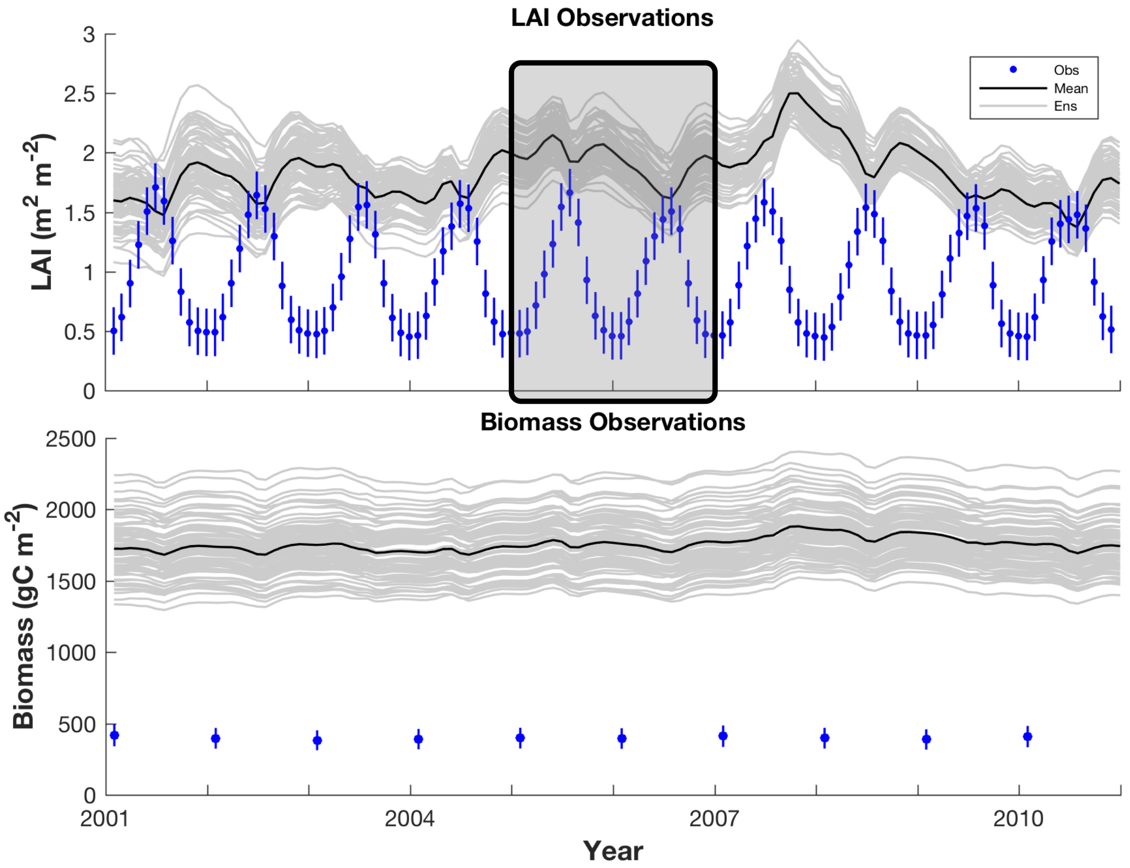
Recent reversal in loss of global terrestrial biomass

Yi Y. Liu[1,2*], Albert I. J. M. van Dijk[3,4], Richard A. M. de Jeu[5], Josep G. Canadell[6], Matthew F. McCabe[7], Jason P. Evans[1] and Guojie Wang[8]

LAI and Biomass – observations

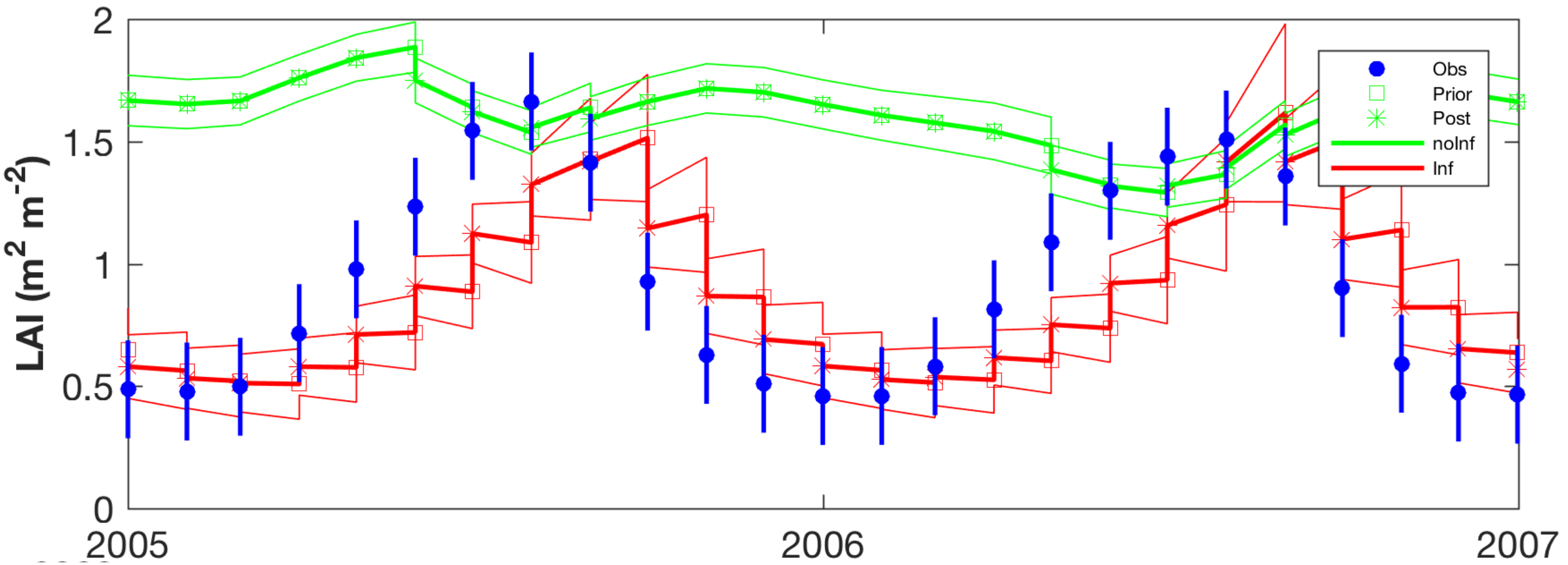


LAI and Biomass – observations



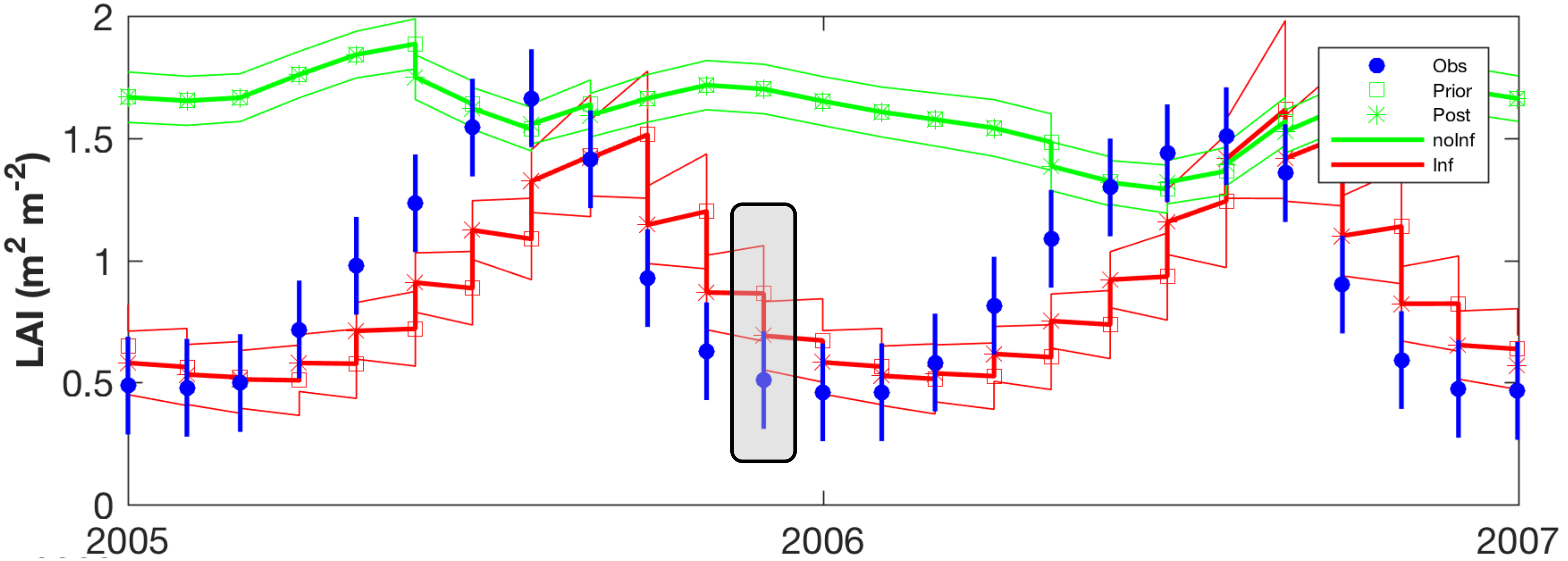
NCAR THE UNIVERSITY OF ARIZONA

Ensemble forecast is updated by observations

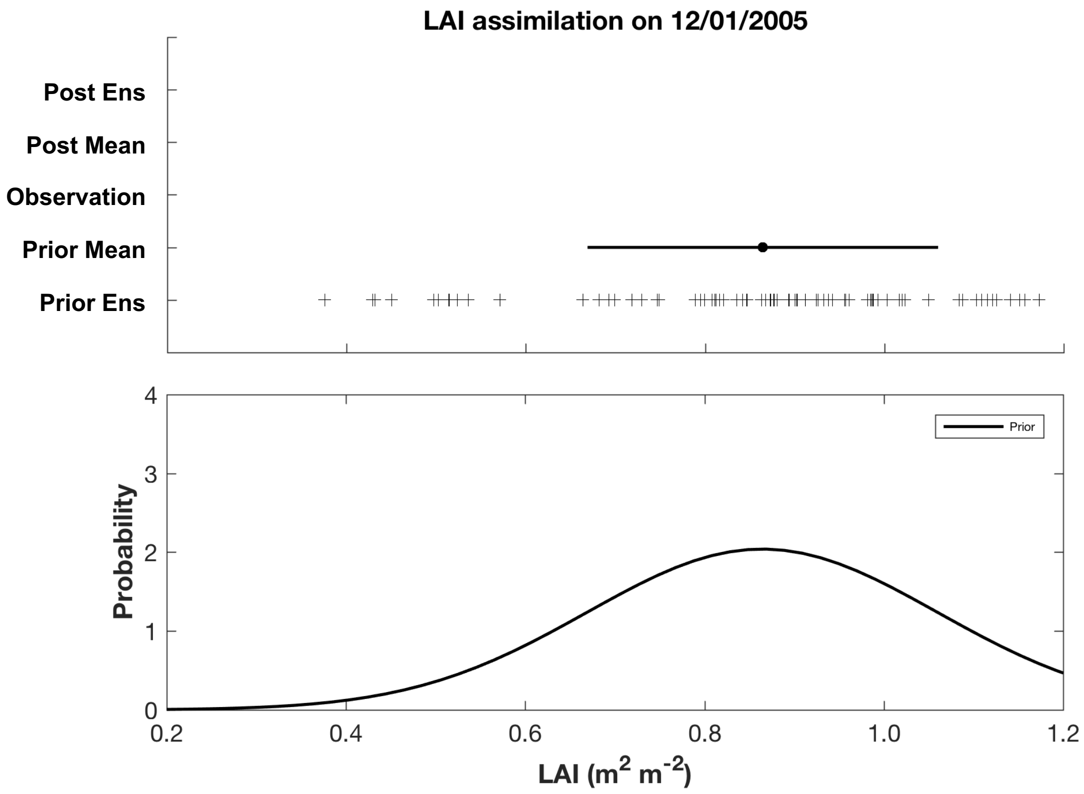


NCAR THE UNIVERSITY OF ARIZONA

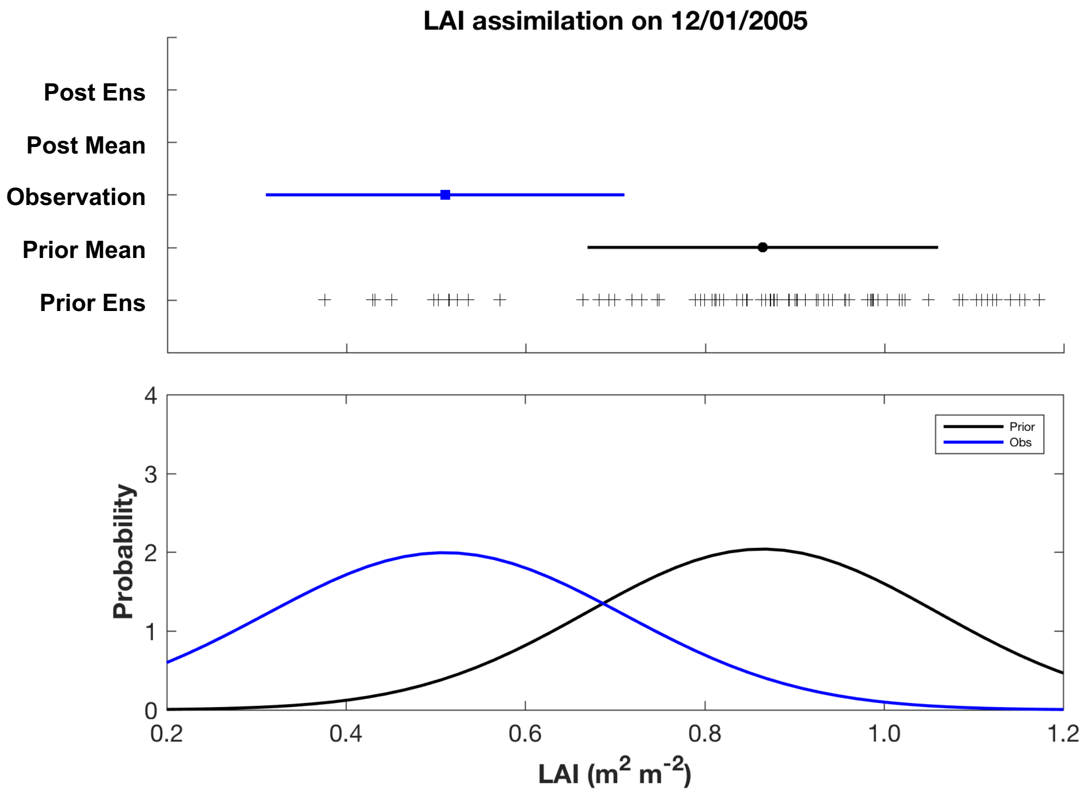
Ensemble forecast is updated by observations



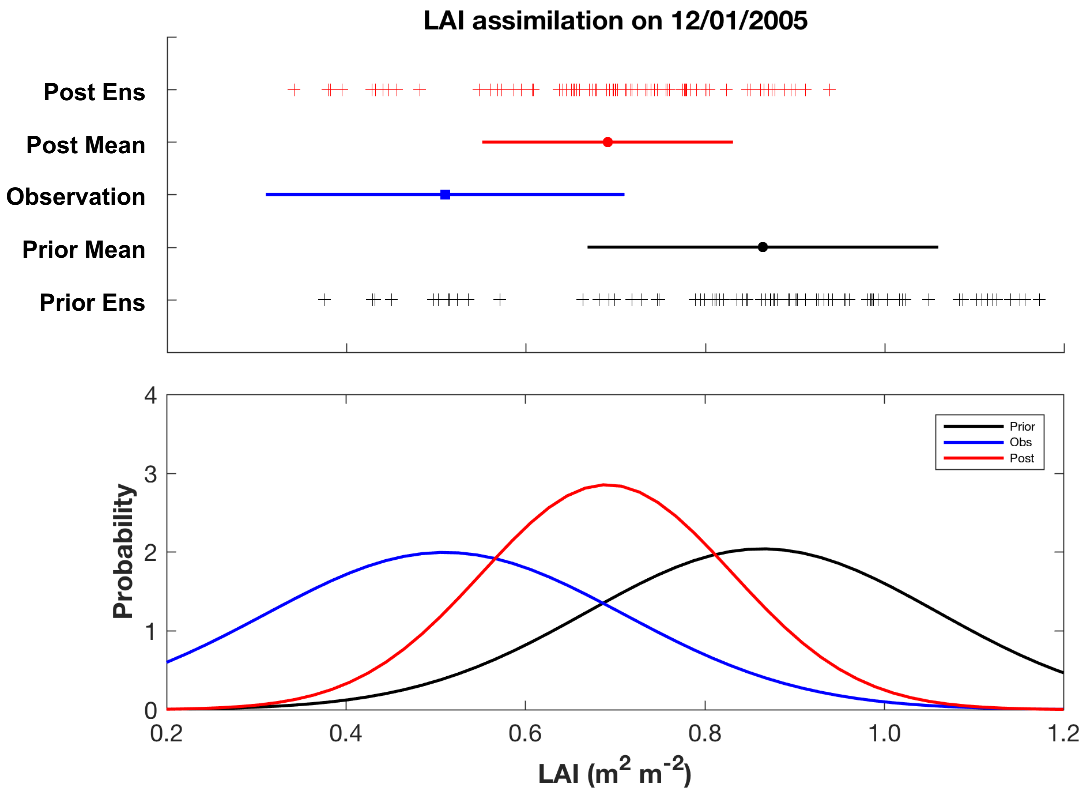
Normal is fitted to the prior/forecast ensemble…



…we have an observation with an uncertainty…

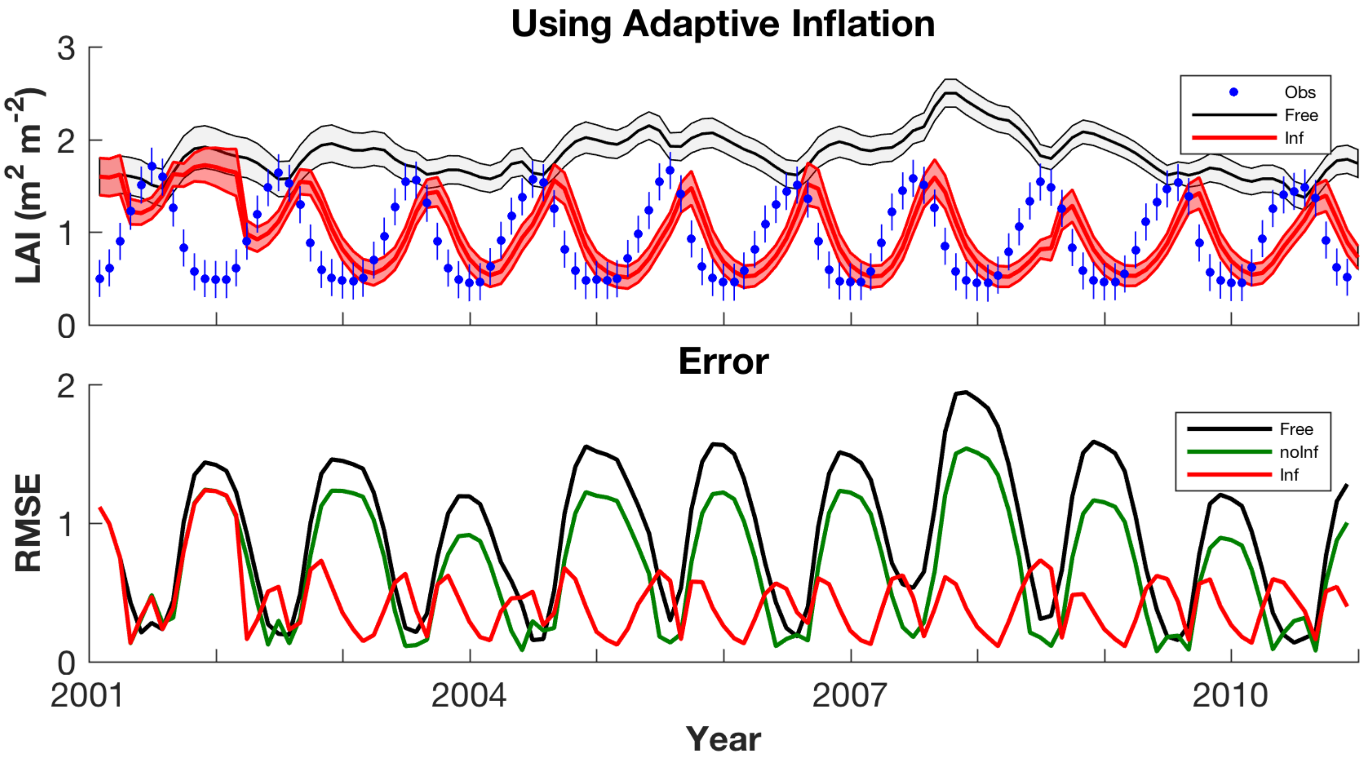


...use EAKF to calculate posterior/analysis



NCAR THE UNIVERSITY OF ARIZONA

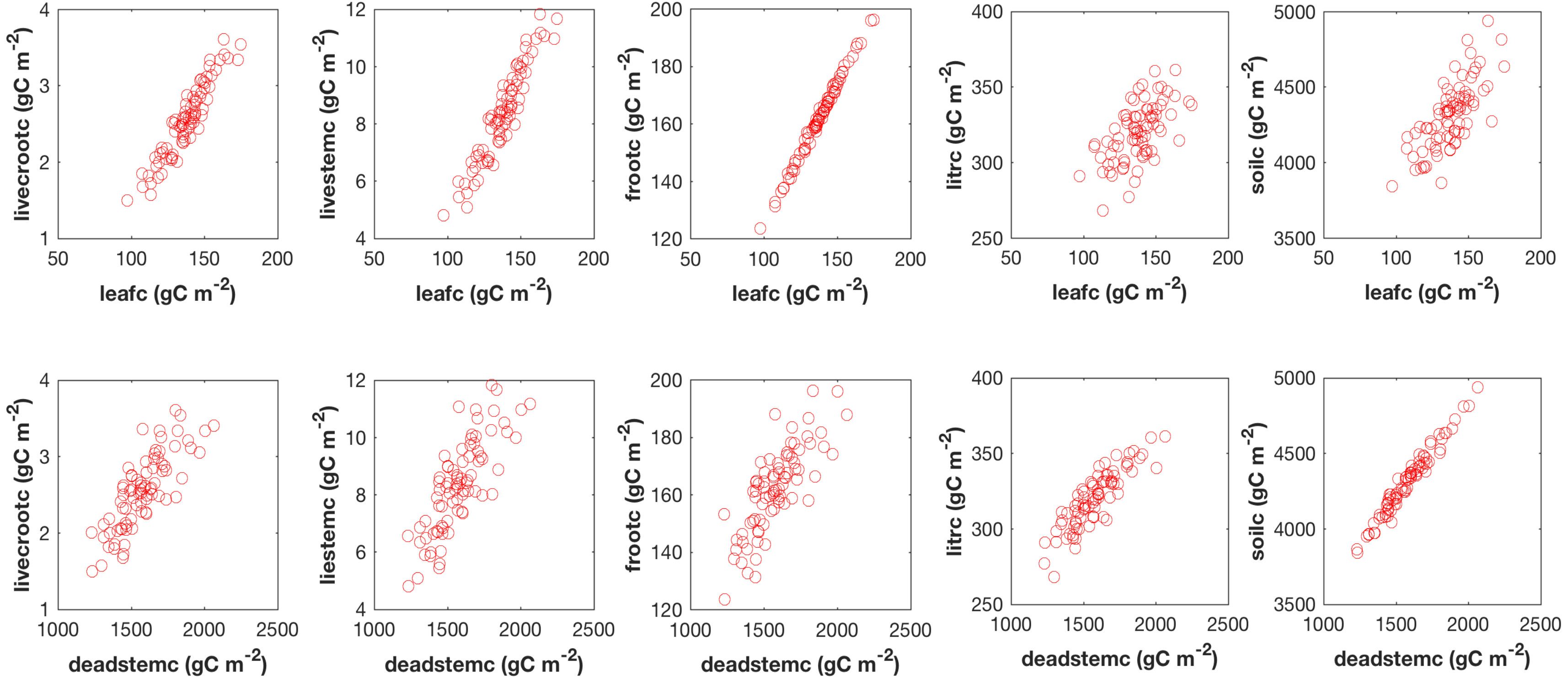
50% reduction in LAI RMSE with assimilation



Exp.	RMSE (m^2 m^{-2})
Free	0.93
Assim	0.44

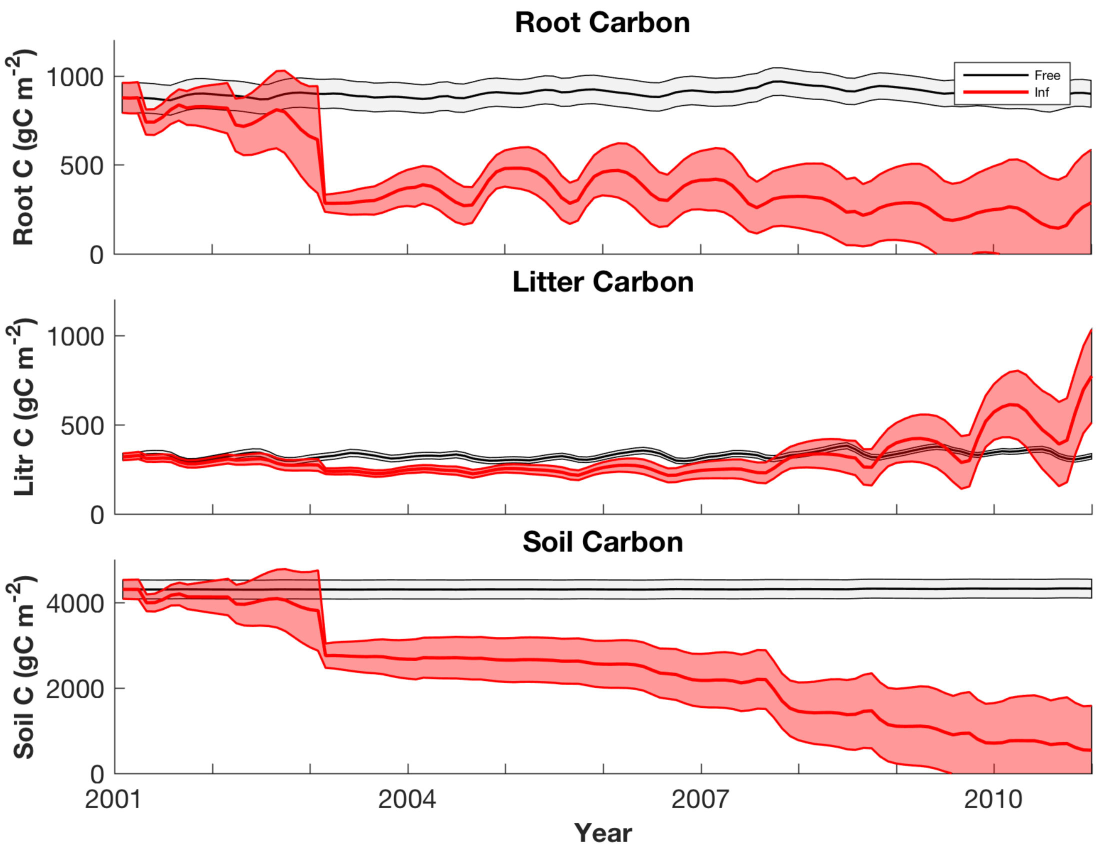
NCAR THE UNIVERSITY OF ARIZONA

Observed and unobserved states



NCAR THE UNIVERSITY OF ARIZONA

Unobserved State variables are also updated



More of the gory details can be found here…

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1029/2018MS001362

Key Points:

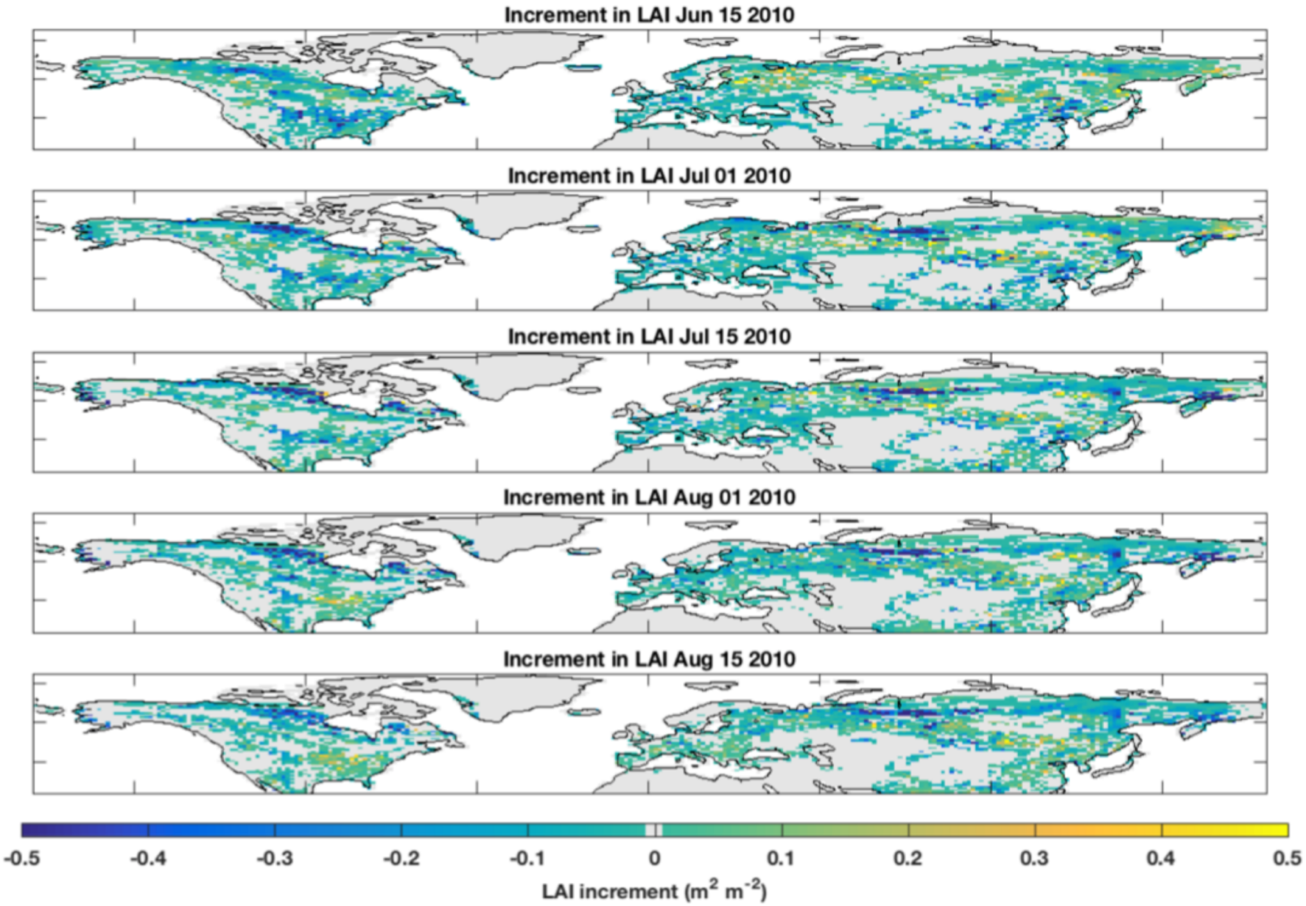
- Data assimilation was used to initialize biomass and leaf area in the Community Land Model
- Adaptive inflation was needed to give more weight to observations due to substantial discrepancies between model forecast and observations

Evaluation of a Data Assimilation System for Land Surface Models Using CLM4.5

Andrew M. Fox[1], Timothy J. Hoar[2], Jeffrey L. Anderson[2], Avelino F. Arellano[3], William K. Smith[1], Marcy E. Litvak[4], Natasha MacBean[1], David S. Schimel[5], and David J. P. Moore[1]

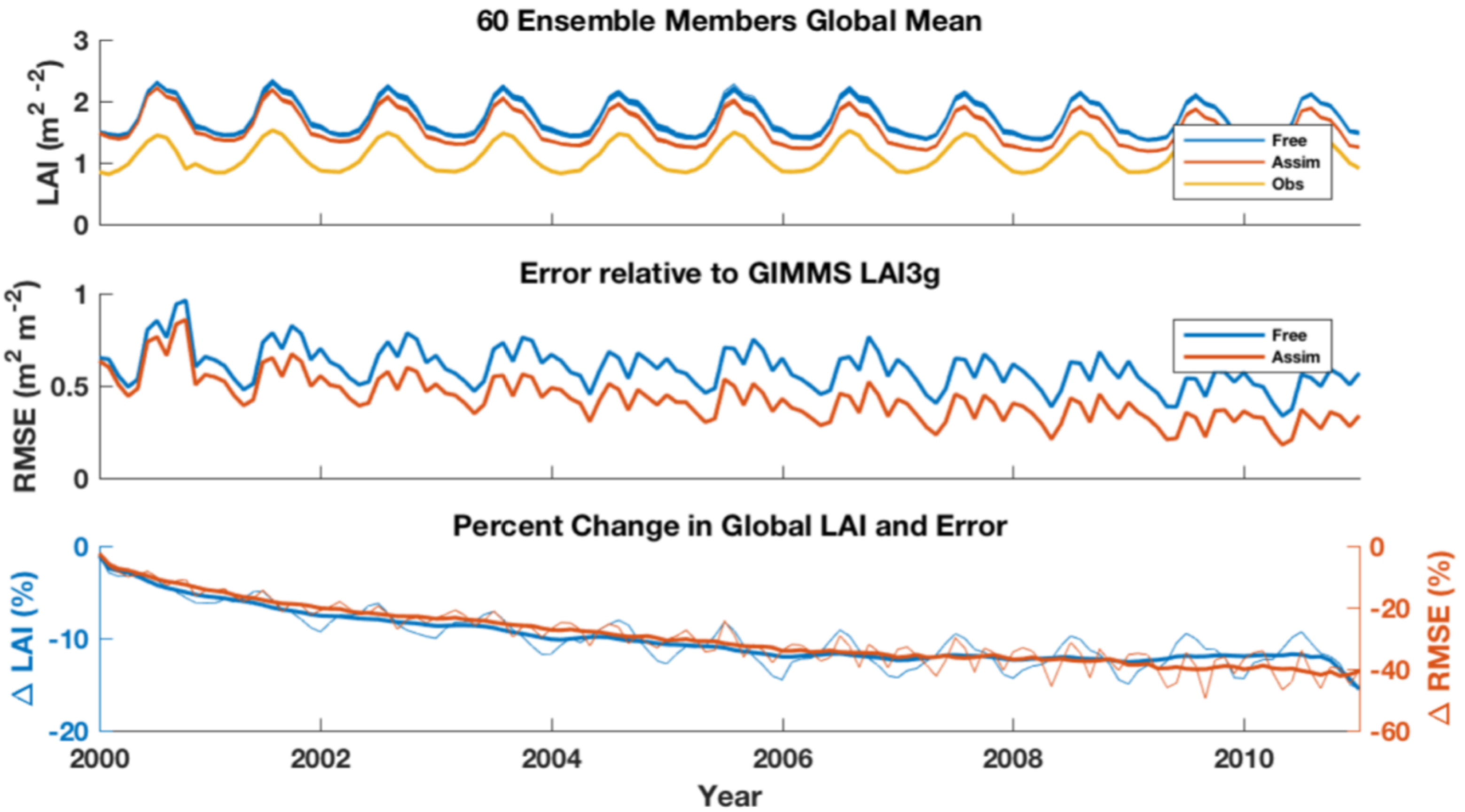
[1]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]Hydrological and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, [4]Department of Biology, University of New Mexico, Albuquerque, NM, USA, [5]Jet Propulsion Laboratory, Pasadena, CA, USA

Updates in LAI calculated every 15 days



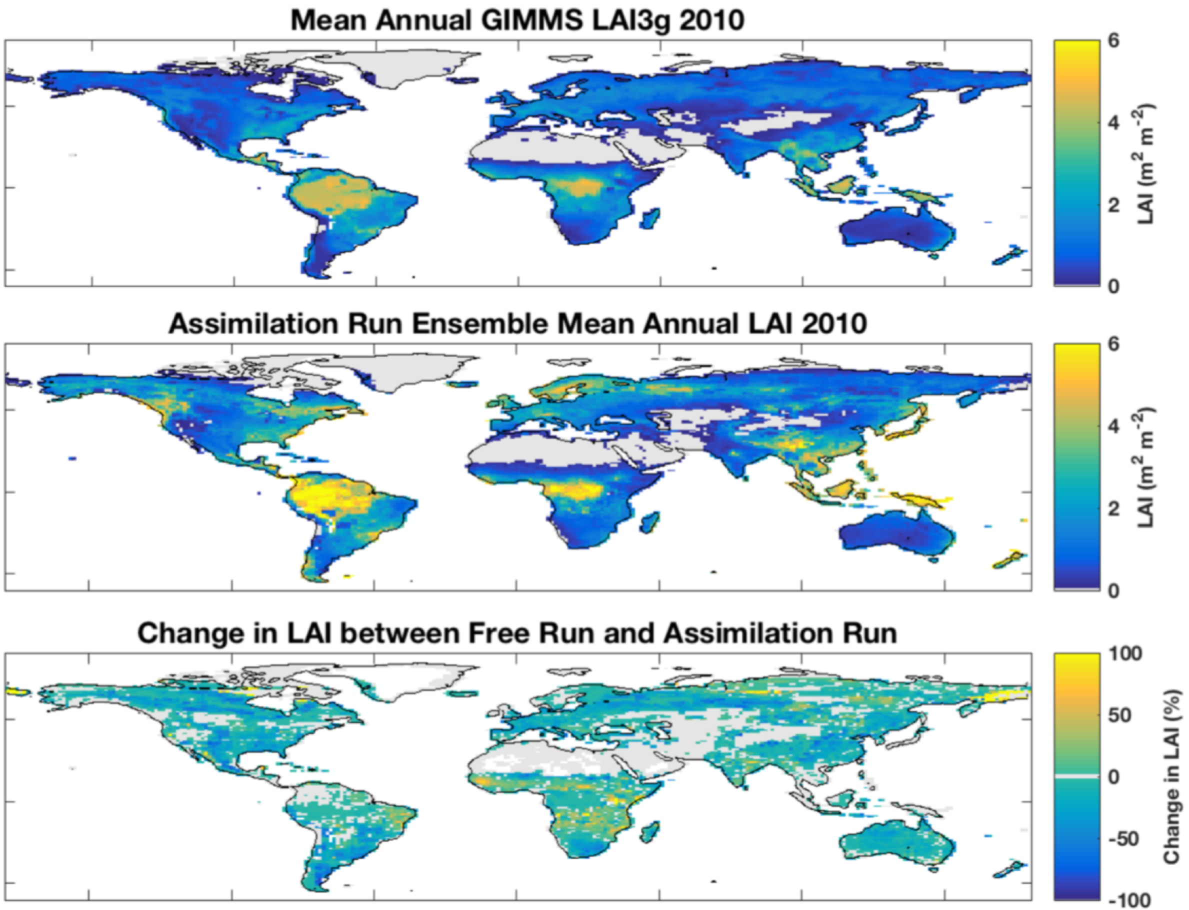
NCAR THE UNIVERSITY OF ARIZONA

Impact on Global LAI



THE UNIVERSITY OF ARIZONA

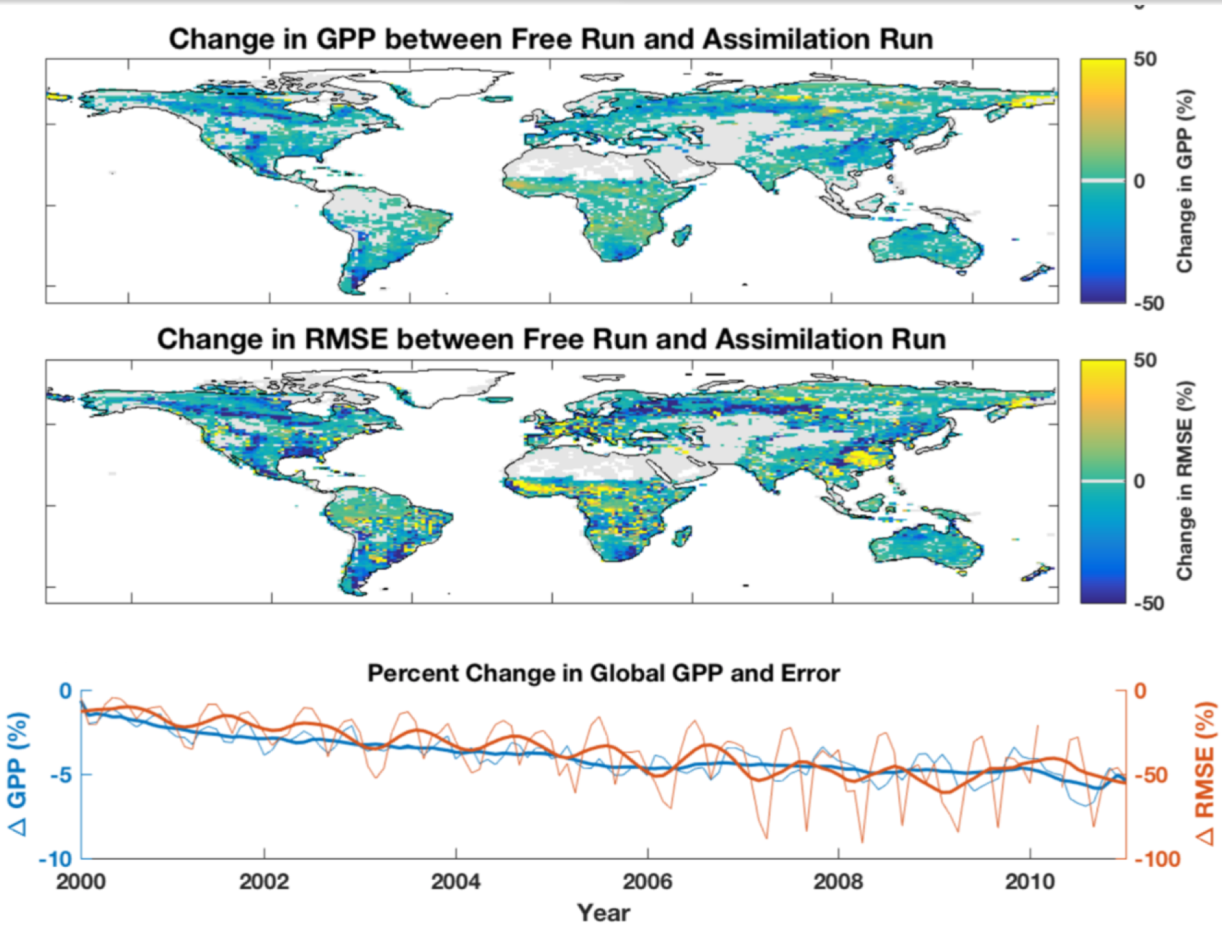
Change in mean annual global LAI



NCAR

The University of Arizona

Impact on Global GPP



HOW WE ACTUALLY DO THIS…

Using CTSM-DART

Basic workflow for CTSM-DART

1) Set up and test a multi-instance case (this has nothing to do with DA *per se*)
2) Prepare some observations
3) Customize the DART assimilation script
4) Enable data assimilation in the case
5) Run the model…

1a. Set up multi-instance case

DART has heavily documented setup scripts to set up a multi-instance global case, *where each instance uses a unique data atmosphere (restart/parameter file)*

./create_newcase --res ${resolution} \
--mach ${machine} \
--compset ${compset} \
--case ${caseroot} \
--project ${project} \
--run-unsupported \
**--ninst ${num_instances} **
--multi-driver

1b. Set up multi-instance case

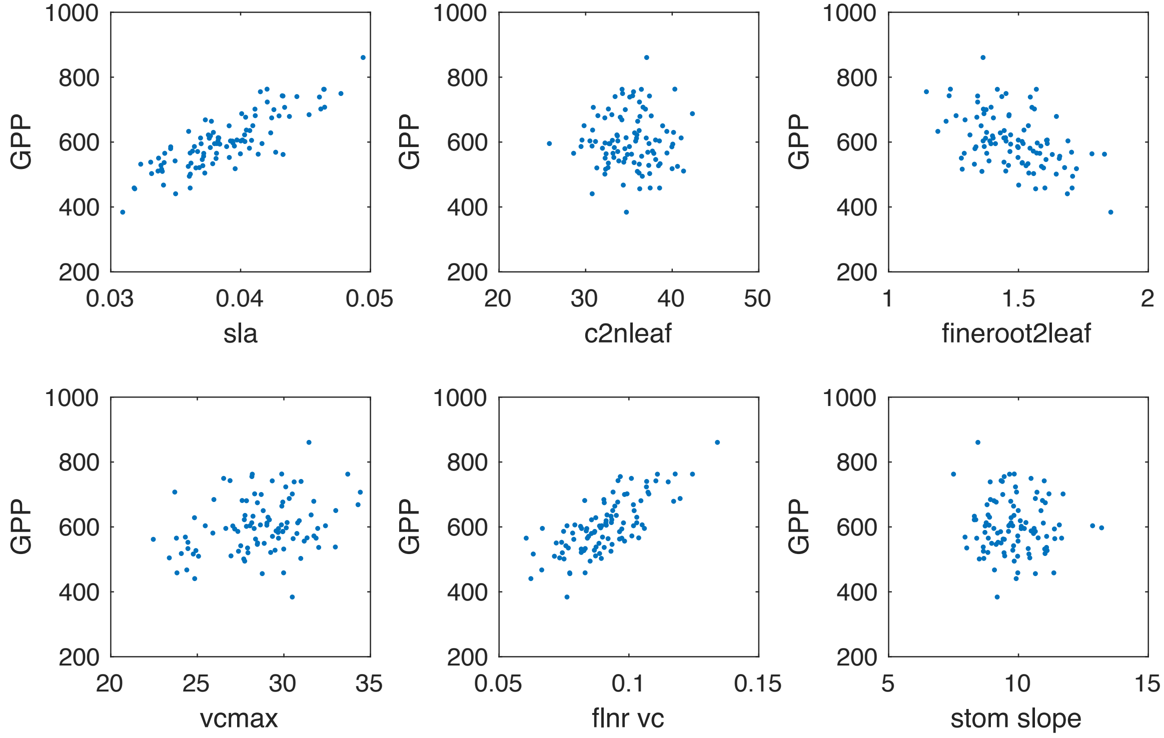
It then runs many **./xmlchange** commands to change PE layout, REFDATE, STARTDATE, etc., etc.

./case.setup

Then it manipulates the many user_nl files to point to incrementally numbered, unique datm.stream files, parameter files and so on

./case.setup

An aside…



Multi-instance capability is a really cool way for carrying out parameter sensitivity analysis…

Here we looking at annual GPP sensitivity to key parameters in C4 grass across 200 ensemble members

2. Prepare some observations

- DART has tools to convert many kinds of observations from their raw formats (netCDF, BUFR,HDF,csv, etc.)
- Its convenient to chunk the observations into ‘assimilation-sized’ files and tag each with the CESM time of the intended restart
- If the model stops at midnight, the filenames could be input_obs.2001-01-06-00000 and have all the observations you want to assimilate at that midnight
- The DART example scripts presume the observations have been staged

Data
Assimilation
Research
Testbed

3. Customize the DART scripts

`CESM2_0_DART_config` script is quite simple and heavily documented.

It lives in the *CASEROOT* directory and is run interactively. It has 4 functions

1. Copies the DART executables to *EXEDIR*
2. Copies the DART run-time resources to *RUNDIR*
3. Copies the DART scripts and namelist to *CASEROOT*
4. Uses **`xmlquery`** and **`xmlchange`** for additional customizations

4. Enable the assimilation

CESM2_0_DART_config script then does the following

1. ./**xmlchange** DATA_ASSIMILATION=TRUE
2. ./**xmlchange** DATA_ASSIMILATION_SCRIPT = ${CASEROOT}/assimilate.csh

assimilate.csh contains information about observation file locations, and controls execution of DART

The CESM run script is configured to invoke the DA script if the model forecast was successful. Automatically. No modifications.

Data Assimilation Research Testbed

5. Run…

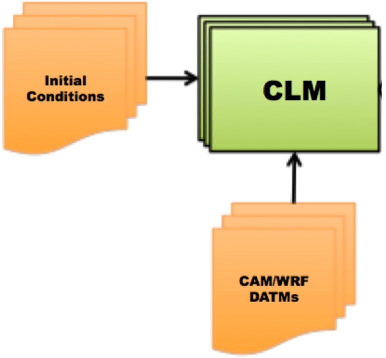
And then

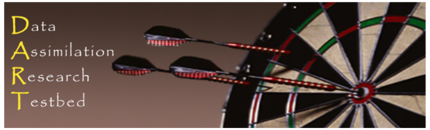
`./case.submit`

As far as the workflow for DA is concerned, *that's it!*

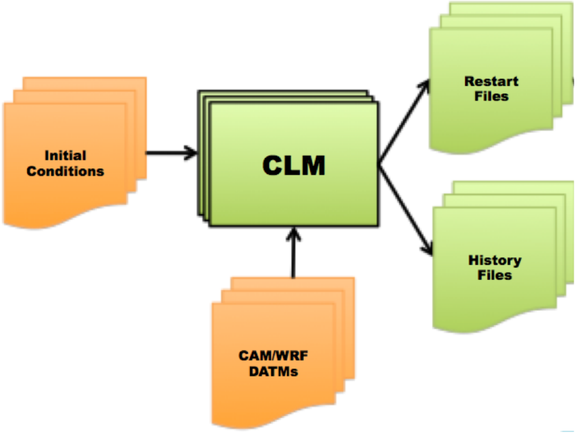
Then you can use the DART diagnostic tools to determine how well the DA system is working…

5. Run…

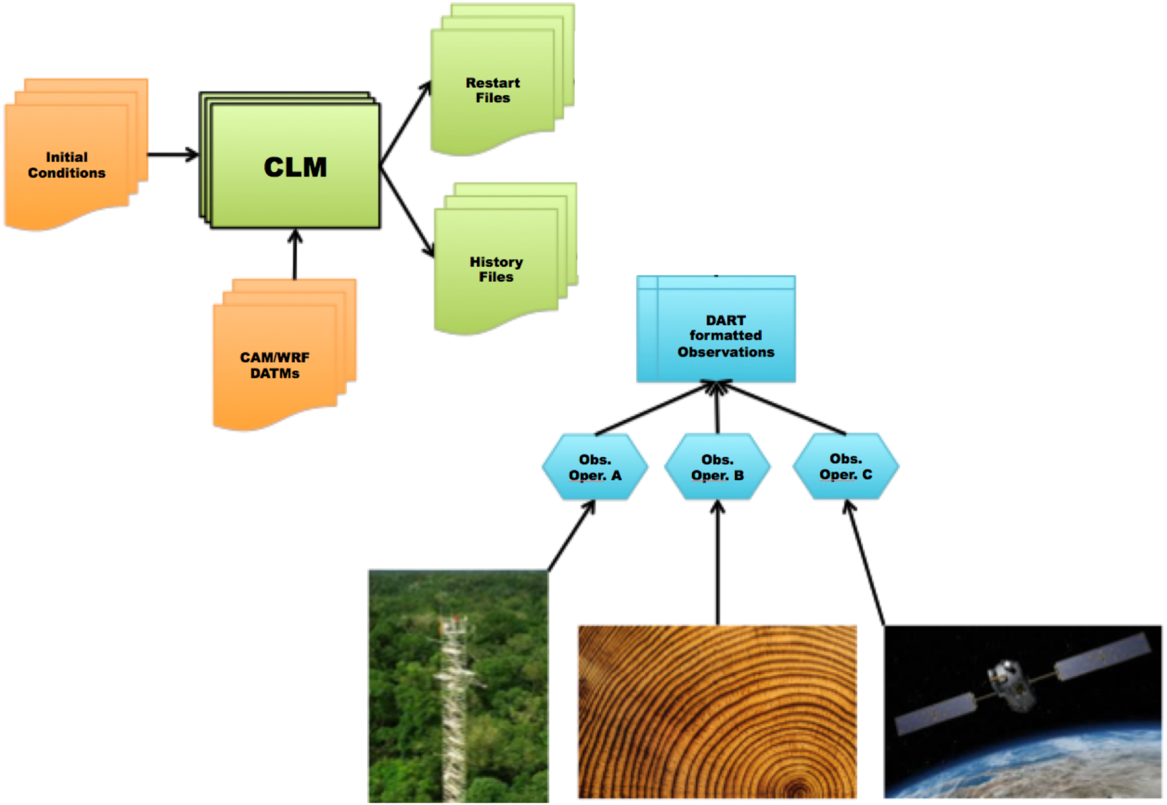

Initial Conditions
CLM
CAM/WRF DATMs


Data
Assimilation
Research
Testbed

5. Run...

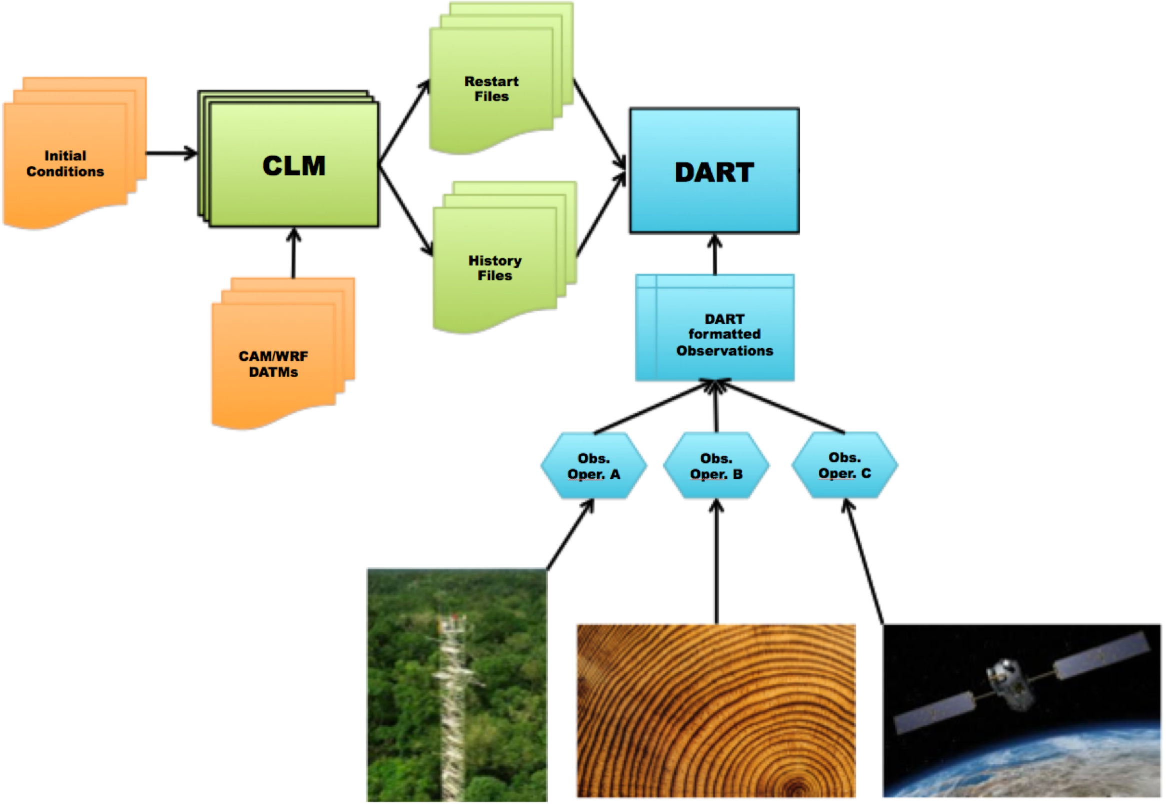

Initial Conditions
CLM
Restart Files
History Files
CAM/WRF DATMs

5. Run…



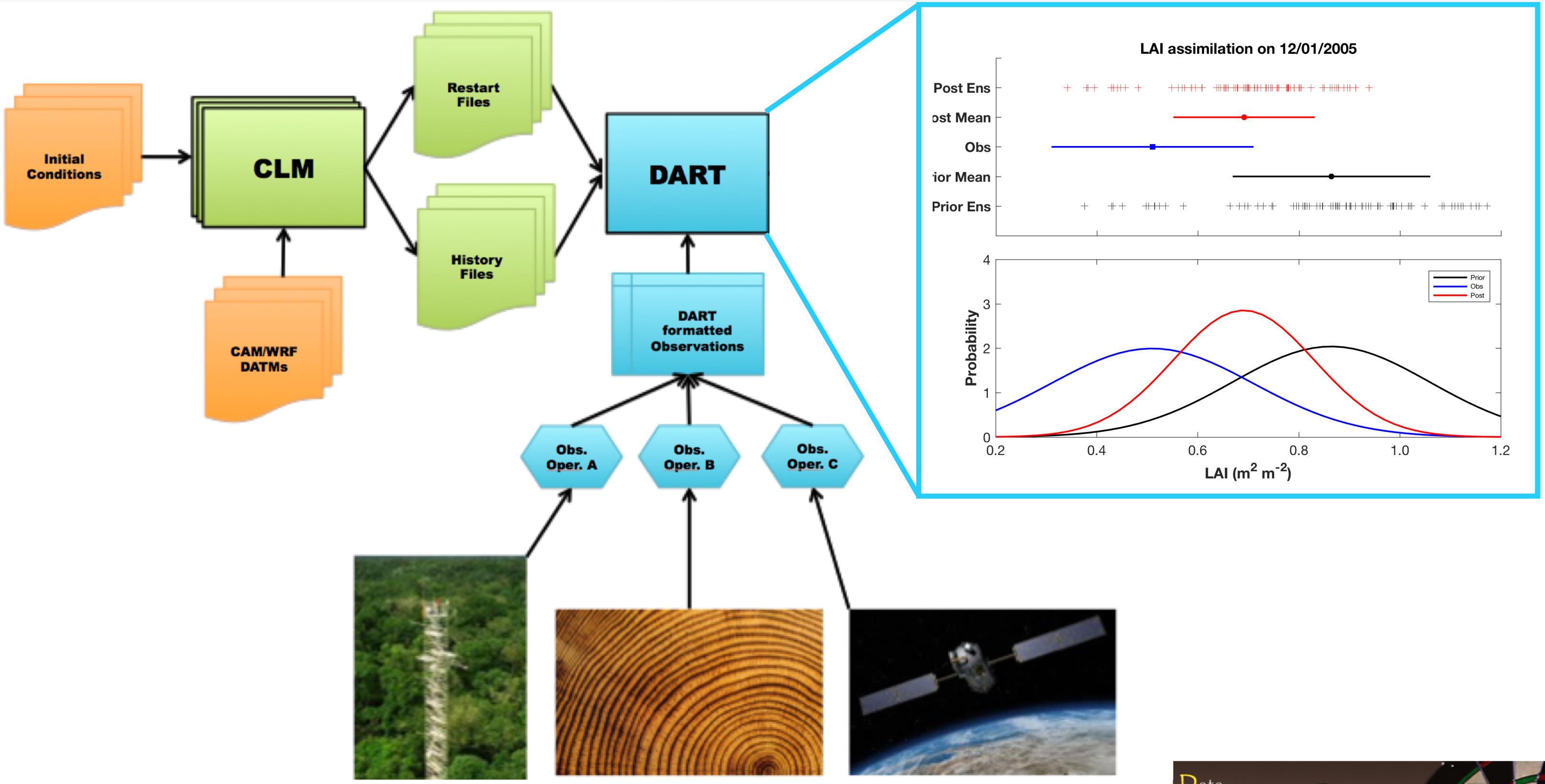
Data
Assimilation
Research
Testbed

5. Run...



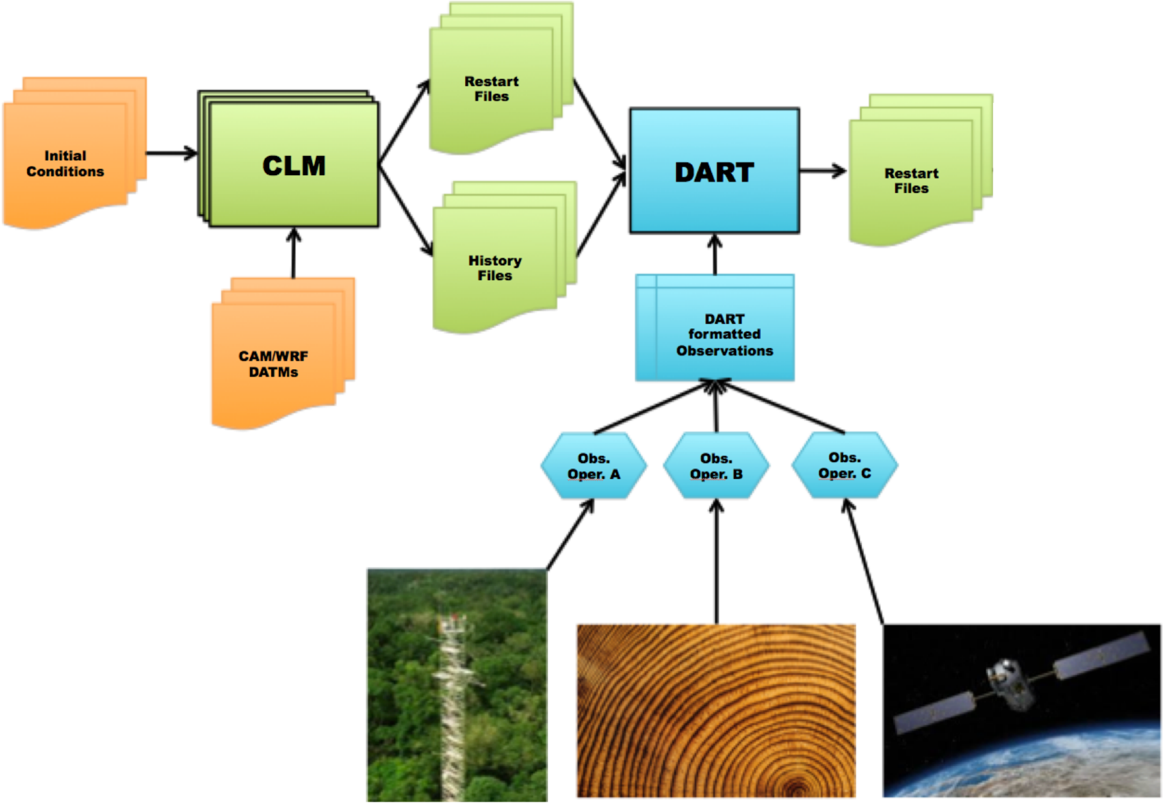
Data Assimilation Research Testbed

5. Run...



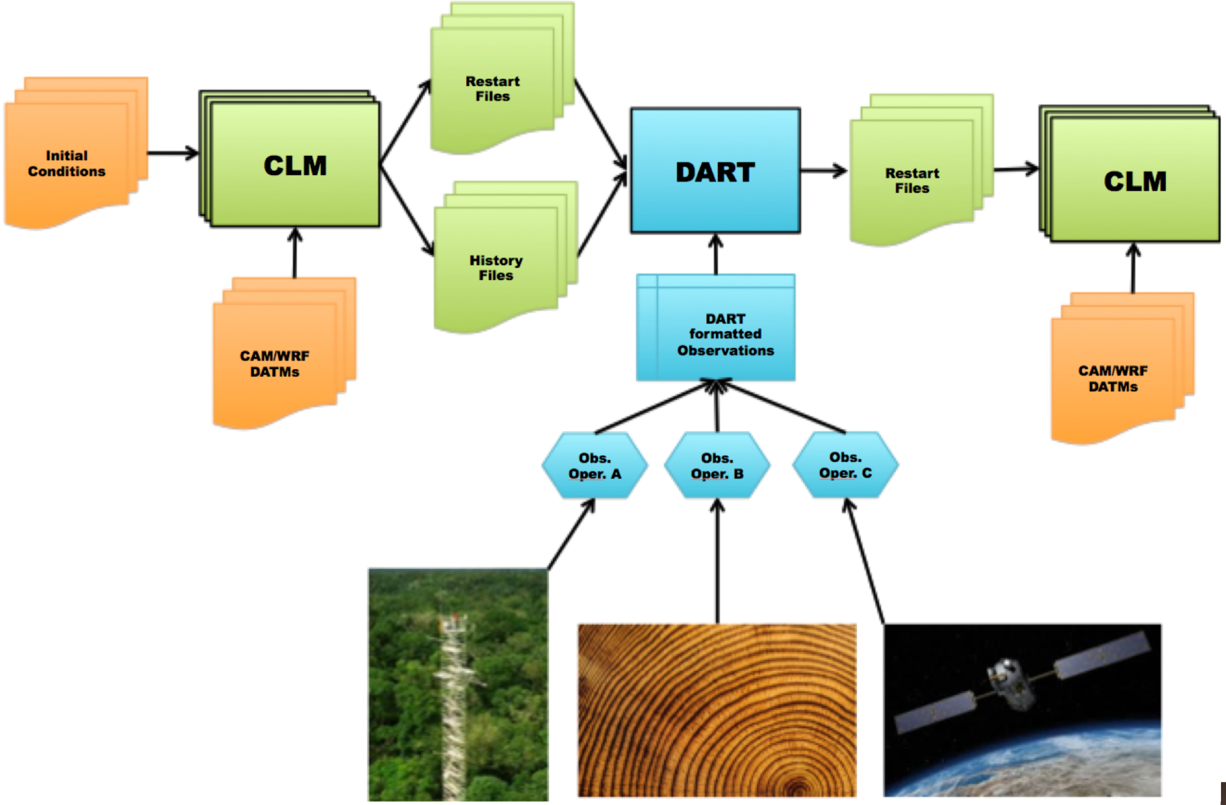
Data
Assimilation
Research
Testbed

5. Run...



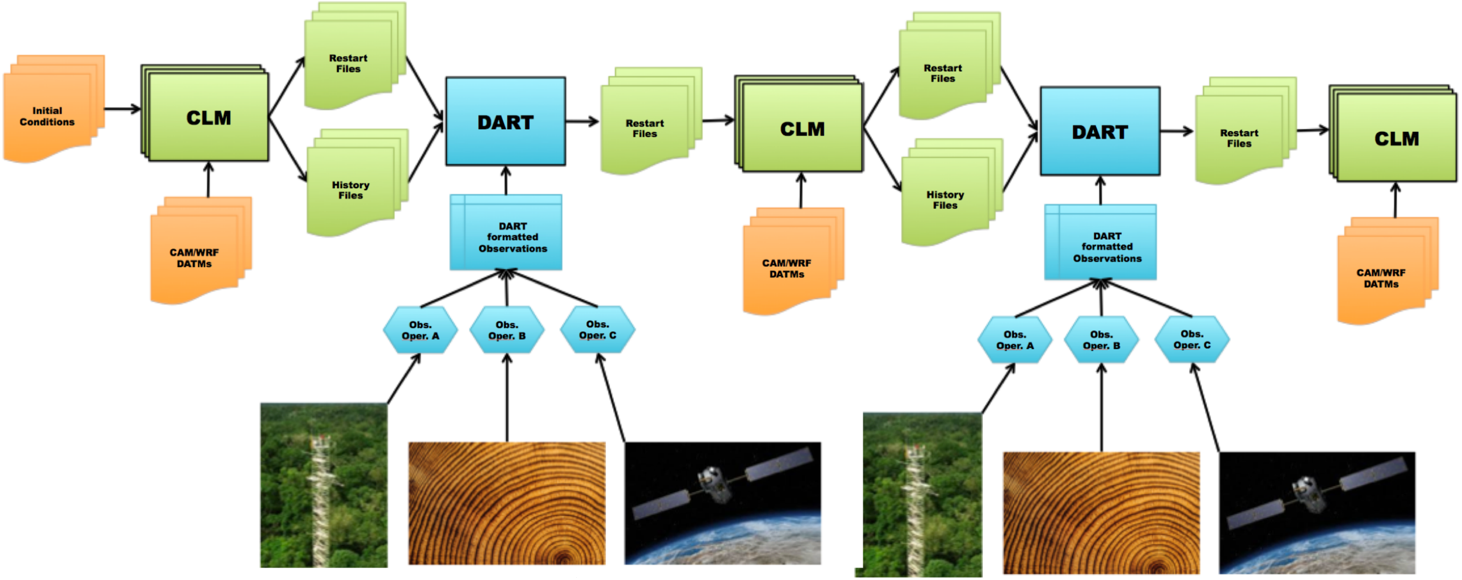
Data
Assimilation
Research
Testbed

5. Run…

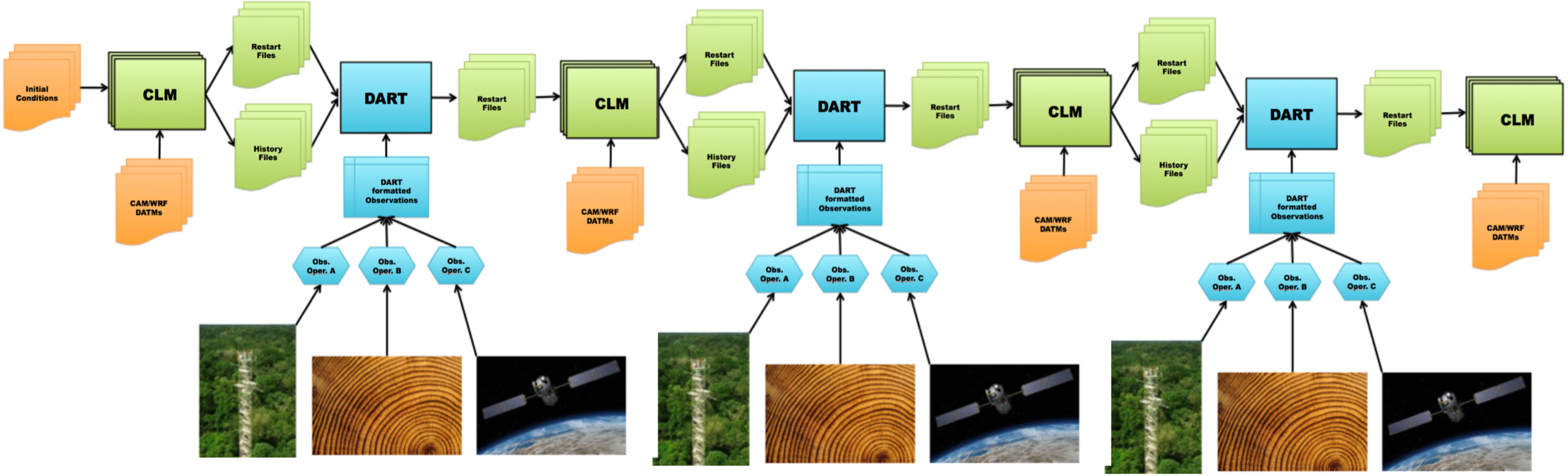


Data Assimilation Research Testbed

5. Run…

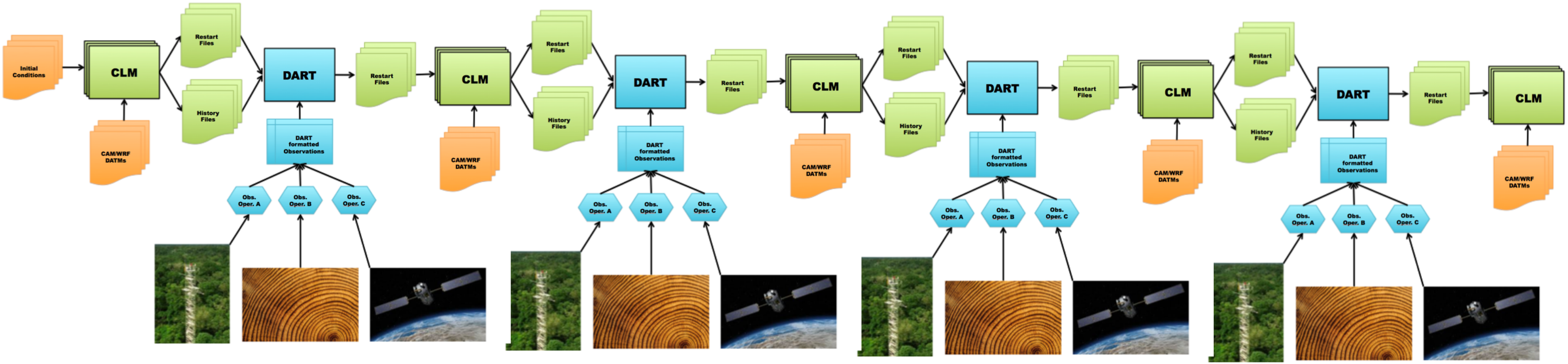

Initial Conditions
CLM
CAM/WRF DATMs
Restart Files
History Files
DART
DART formatted Observations
Obs. Oper. A
Obs. Oper. B
Obs. Oper. C
Restart Files
CLM
CAM/WRF DATMs
Restart Files
History Files
DART
DART formatted Observations
Obs. Oper. A
Obs. Oper. B
Obs. Oper. C
Restart Files
CLM
CAM/WRF DATMs

5. Run…



Data Assimilation Research Testbed

5. Run...



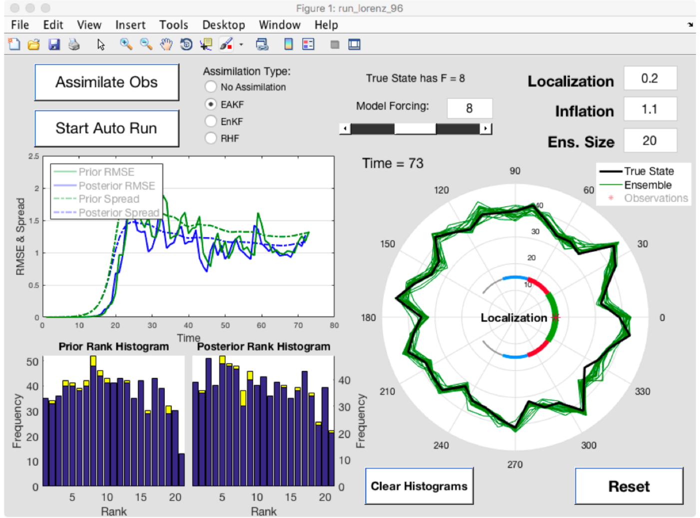
Data Assimilation Research Testbed

The DART Tutorial & DART_LAB

The **DART Tutorial** is a step-by-step approach to ensemble DA.

Section 1	[pdf]	Filtering For a One Variable System
Section 2	[pdf]	The DART Directory Tree
Section 3	[pdf]	DART Runtime Control and Documentation
Section 4	[pdf]	How should observations of a state variable impact an unobserved
Section 5	[pdf]	Comprehensive Filtering Theory: Non-Identity Observations and th
Section 6	[pdf]	Other Updates for An Observed Variable
Section 7	[pdf]	Some Additional Low-Order Models
Section 8	[pdf]	Dealing with Sampling Error
Section 9	[pdf]	More on Dealing with Error; Inflation
Section 10	[pdf]	Regression and Nonlinear Effects
Section 11	[pdf]	Creating DART Executables
Section 12	[pdf]	Adaptive Inflation
Section 13	[pdf]	Hierarchical Group Filters and Localization
Section 14	[pdf]	Observation Quality Control
Section 15	[pdf]	DART Experiments: Control and Design
Section 16	[pdf]	Diagnostic Output
Section 17	[pdf]	Creating Observation Sequences
Section 18	[pdf]	Lost in Phase Space: The Challenge of Not Knowing the Truth
Section 19	[pdf]	DART-Compliant Models and Making Models Compliant: Coming Sc
Section 20	[pdf]	Model Parameter Estimation
Section 21	[pdf]	Observation Types and Observing System Design
Section 22	[pdf]	Parallel Algorithm Implementation: Coming Soon
Section 23	[pdf]	Location Module Design
Section 24		Fixed Lag Smoother (not available yet)
Section 25	[pdf]	A Simple 1D Advection Model: Tracer Data Assimilation

DART_LAB is a set of PDF presentation files and a set of MATLAB® examples that comprise a fully self-contained introduction to Data Assimilation and the Ensemble DA concepts.



www.image.ucar.edu/DAReS/DART/

Email: dart@ucar.edu

NCAR THE UNIVERSITY OF ARIZONA

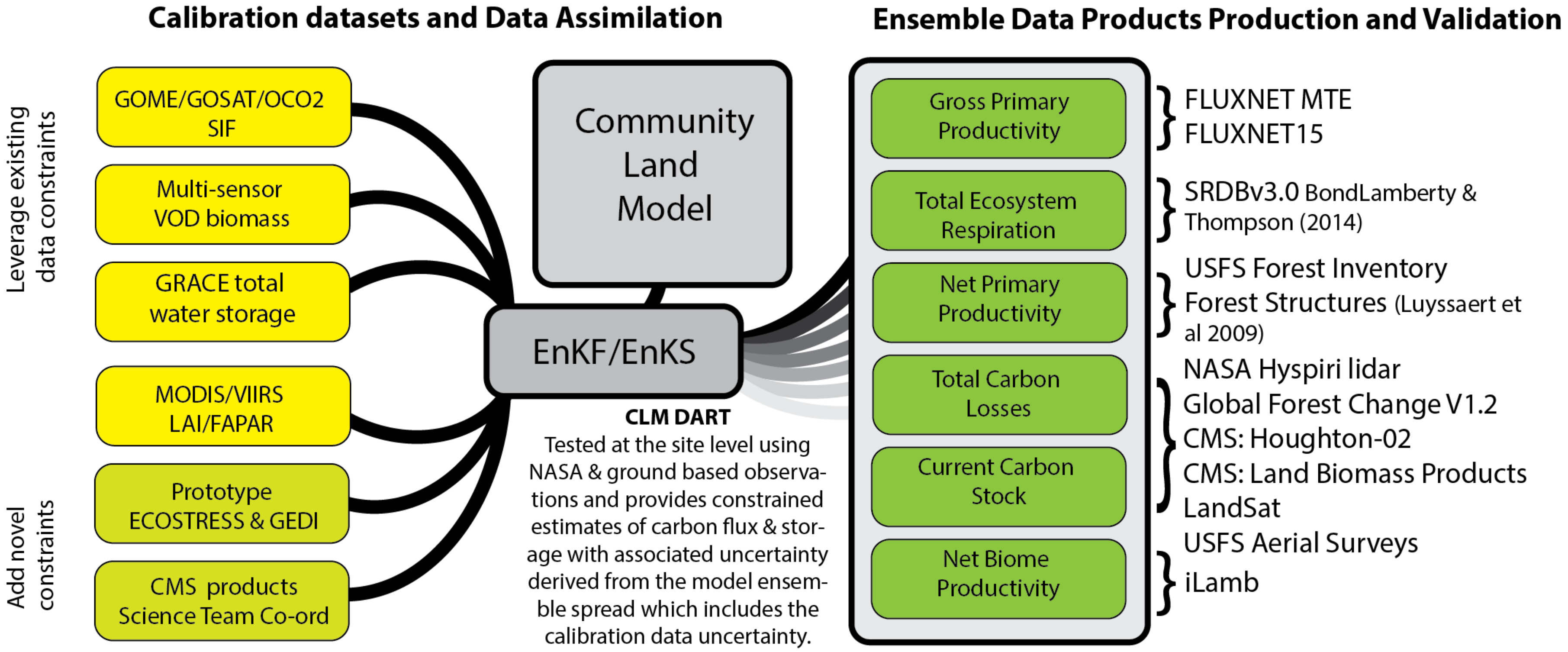
WHAT ARE WE THINKING ABOUT NEXT?

Future directions

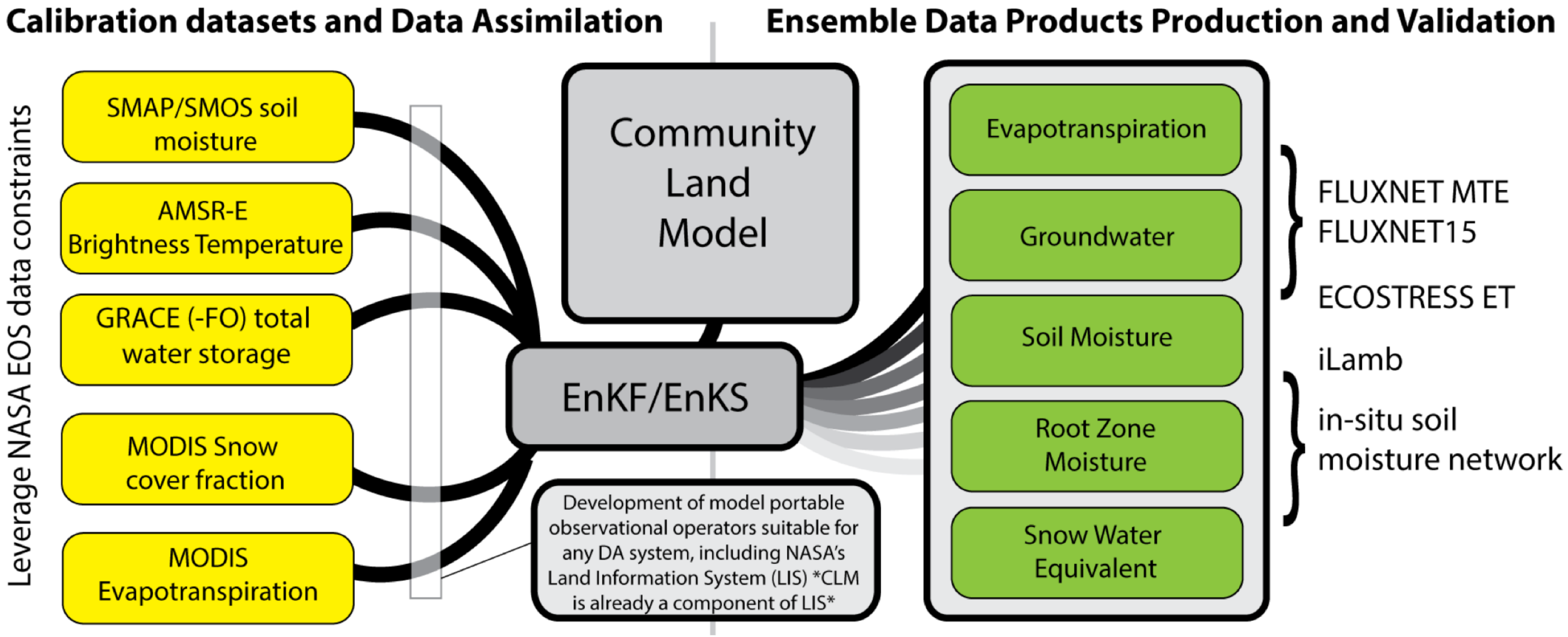
Future Directions

1) Merging multiple types of RS observations

Merging RS data and models - Carbon



Merging RS data and models – Water



NCAR THE UNIVERSITY OF ARIZONA

Future Directions

1) Merging multiple types of RS observations
2) Working with new observations

NCAR THE UNIVERSITY OF ARIZONA

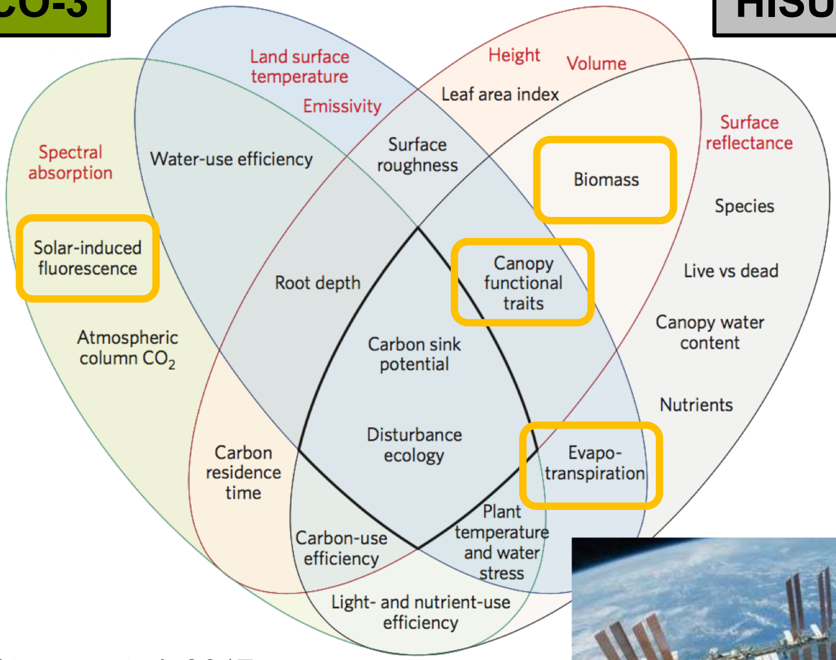
New observations from the ISS

Early 2019 **OCO-3**

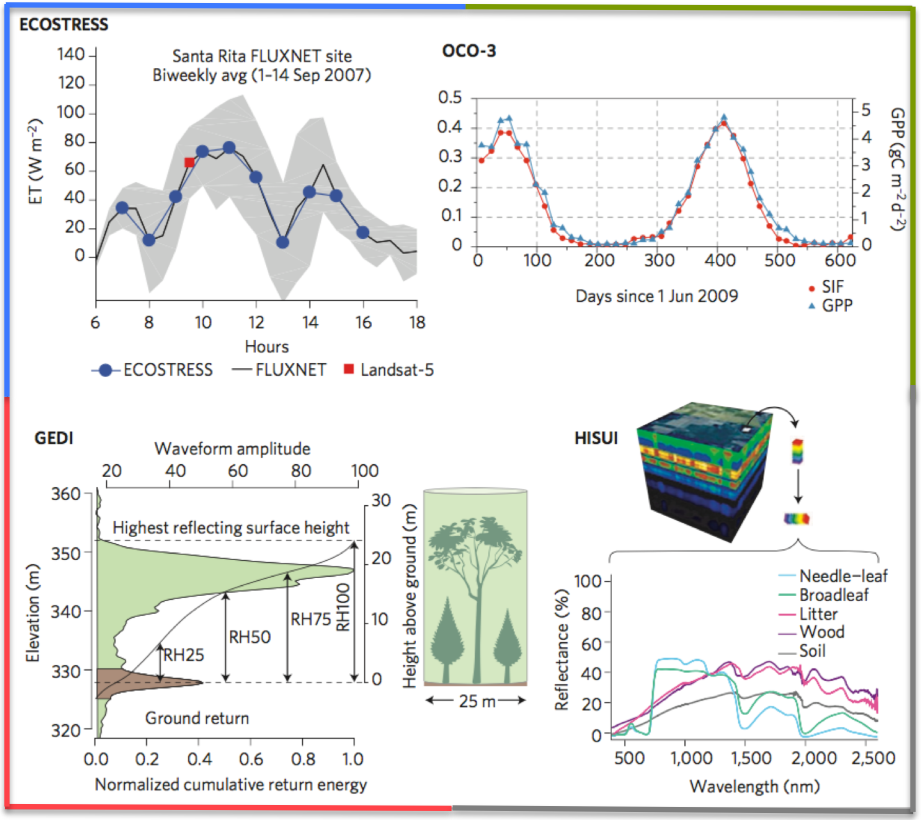
July 2018 **ECOSTRESS**

Dec 2018 **GEDI**

Early 2020 **HISUI**

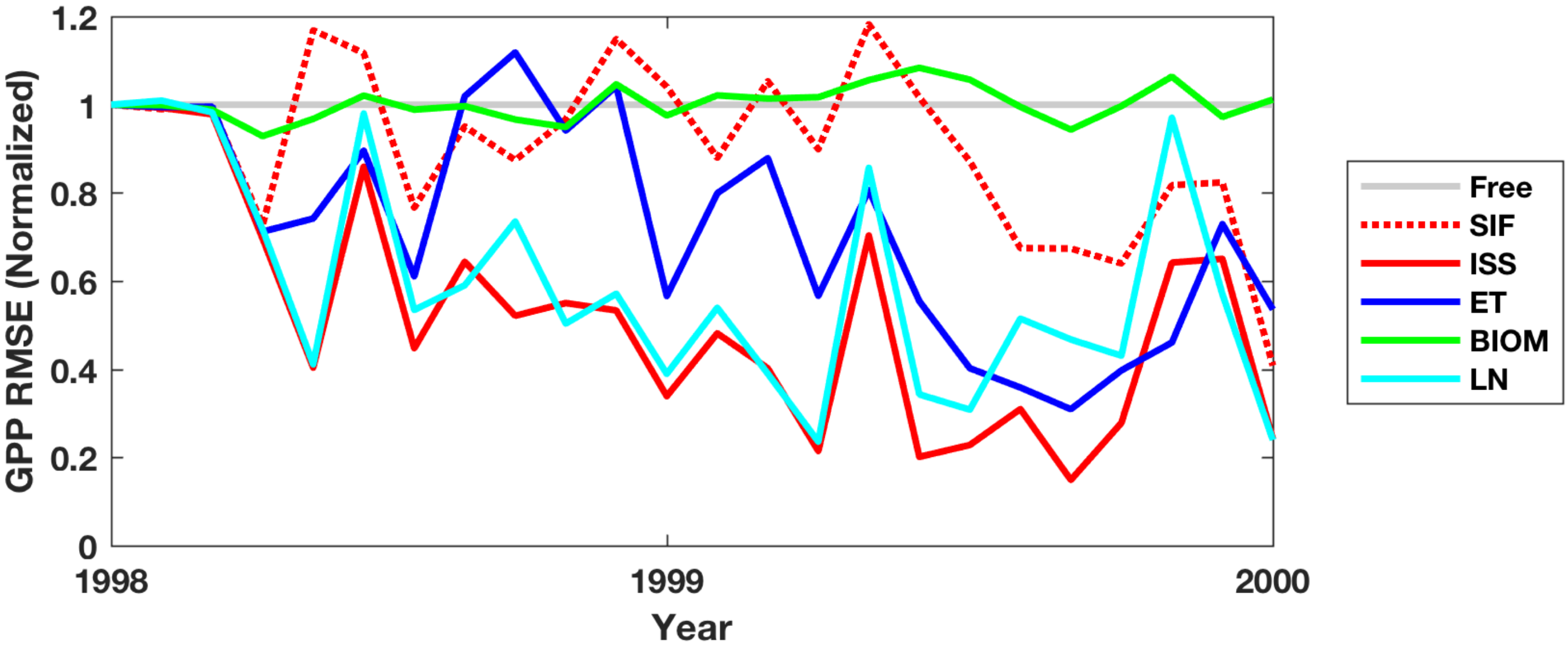


Stavros et al. 2017



NCAR THE UNIVERSITY OF ARIZONA

Combining all ISS obs reduces error the most



NCAR THE UNIVERSITY OF ARIZONA

Future Directions

1) Merging multiple types of RS observations
2) Working with new observations
3) Moving from data products to “raw observations”

ECOSTRESS Level-3 Evapotranspiration ATBD

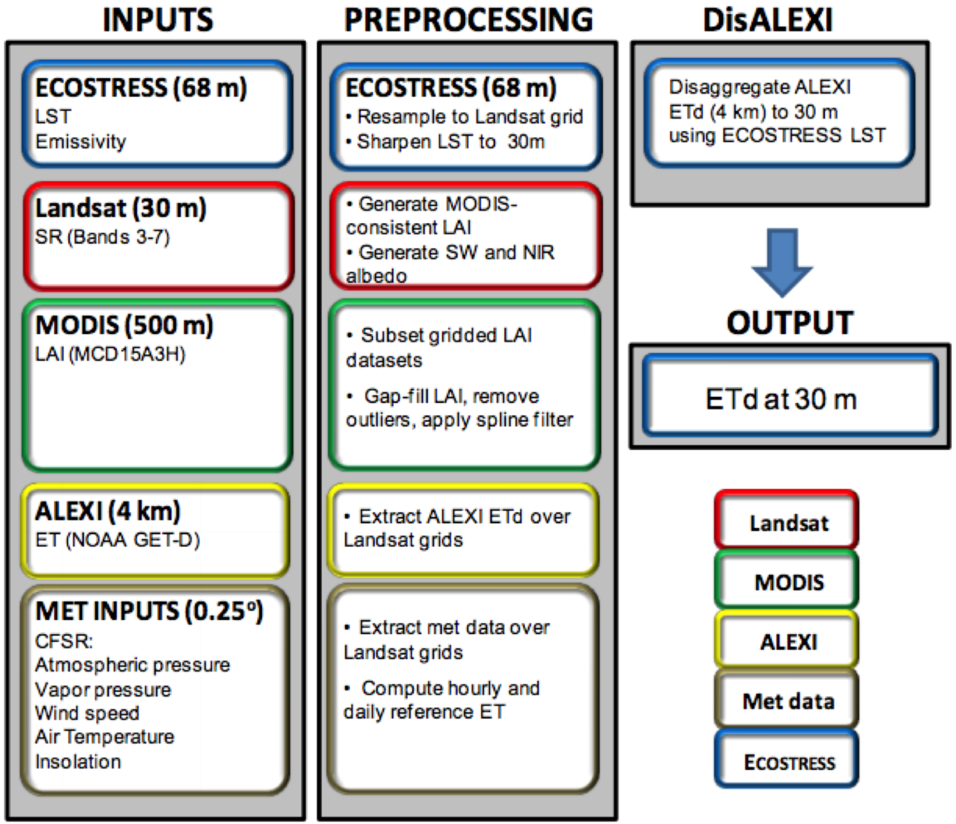
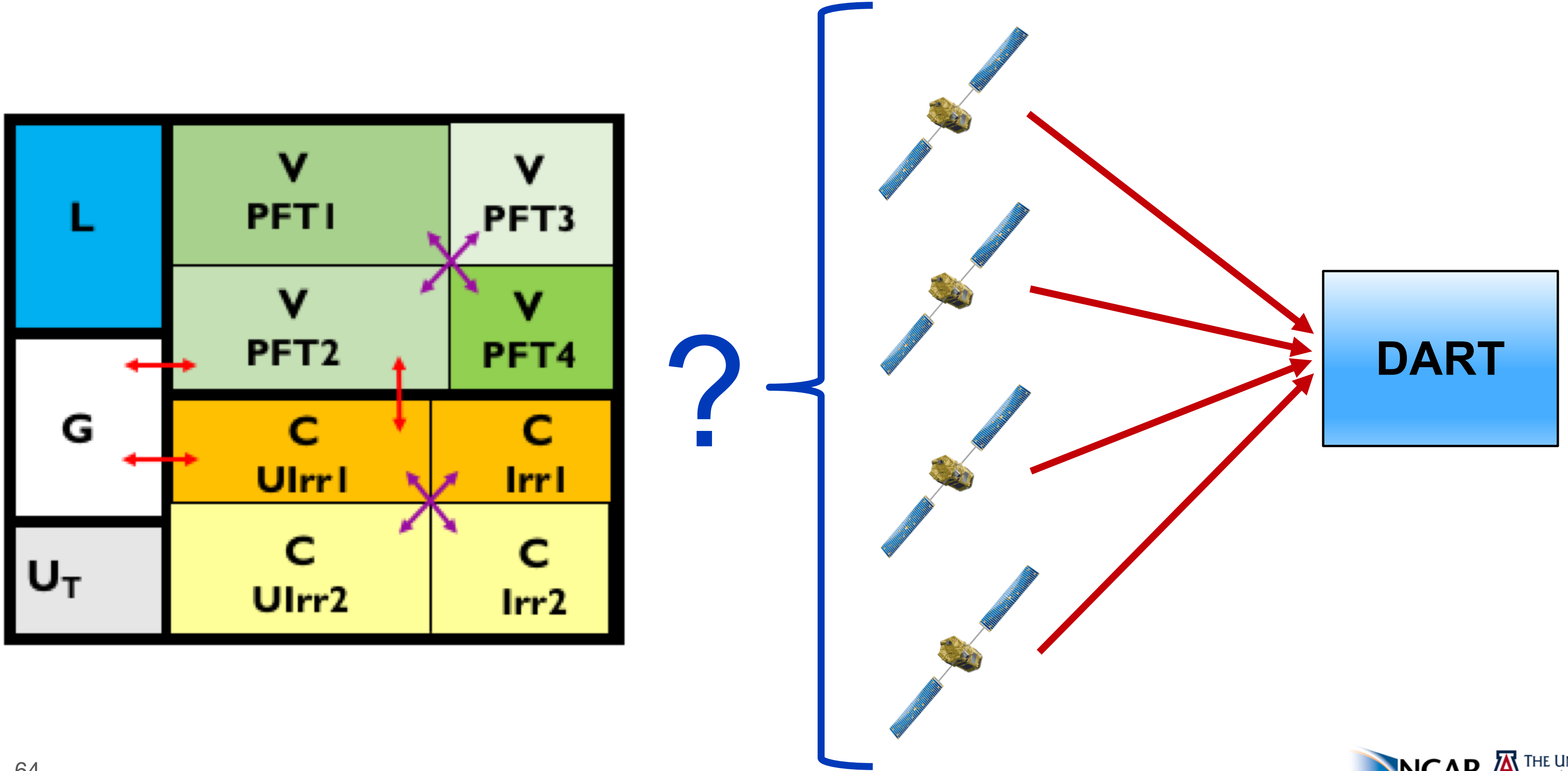


Figure 4. Conceptual diagram describing computation of L-3(ALEXI_ET) evapotranspiration.

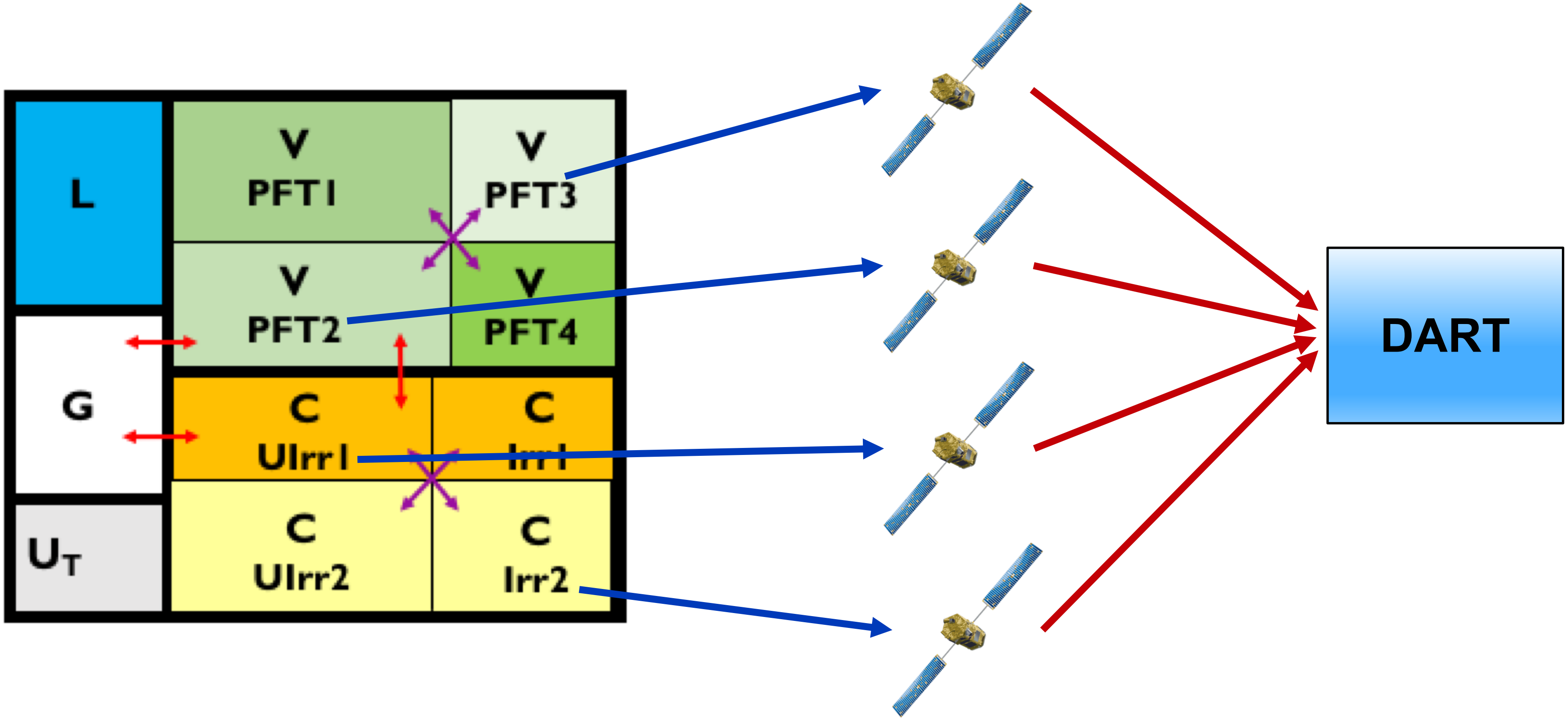
Future Directions

1) Merging multiple types of RS observations
2) Working with new observations
3) Moving from data products to "raw observations"
4) Treating observations as PFT-specific

From grid cell "average" observations to PFT-specific



From grid cell “average” observations to PFT-specific



Future Directions

1) Merging multiple types of RS observations
2) Working with new observations
3) Moving from data products to "raw observations"
4) Treating observations as PFT-specific
5) "Parameter" estimation

   (Often more like special state variables...)

Parameter estimation with EAKF

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE

10.1002/2015JG003297

Key Points:

- The CLM parameters, estimated separately for four plant functional types, correlated with initial carbon-nitrogen pools

Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites

Hanna Post[1,2,3], Jasper A. Vrugt[2,4,5], Andrew Fox[6], Harry Vereecken[2,3], and Harrie-Jan Hendricks Franssen[2,3]

MCMC

Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Advances in Water Resources 28 (2005) 135–147

Advances in Water Resources

www.elsevier.com/locate/advwatres

Dual state–parameter estimation of hydrological models using ensemble Kalman filter

Hamid Moradkhani [a,*], Soroosh Sorooshian [a], Hoshin V. Gupta [b], Paul R. Houser [c]

JOURNAL OF ADVANCES IN MODELING EARTH SYSTEMS, VOL. 5, 58–70, doi:10.1029/2012MS000167, 2013

Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world

Sebastian Schirber,[1] Daniel Klocke,[2] Robert Pincus,[3] Johannes Quaas,[4] and Jeffrey L. Anderson[5]

MONTHLY WEATHER REVIEW VOLUME 143

Parameter Estimation Using Ensemble-Based Data Assimilation in the Presence of Model Error

JUAN RUIZ

Centro de Investigaciones del Mar y la Atmósfera (CIMA/CONICET-UBA), DCAO/FCEyN-Universidad de Buenos Aires, UMI-IFAECI/CNRS, Buenos Aires, Argentina, and AICS/RIKEN, Kobe, Japan

Check for updates

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2017MS001222

Key Points:

The ensemble data assimilation method can potentially be used to

Estimating Convection Parameters in the GFDL CM2.1 Model Using Ensemble Data Assimilation

Shan Li[1,2], Shaoqing Zhang[3,4], Zhengyu Liu[5], Lv Lu[6], Jiang Zhu[2], Xuefeng Zhang[7], Xinrong Wu[7], Ming Zhao[8], Gabriel A. Vecchi[9], Rong-Hua Zhang[4,10], and Xiaopei Lin[3,4]

NCAR THE UNIVERSITY OF ARIZONA

Georgia O’Keeffe – “Black Mesa Landscape”