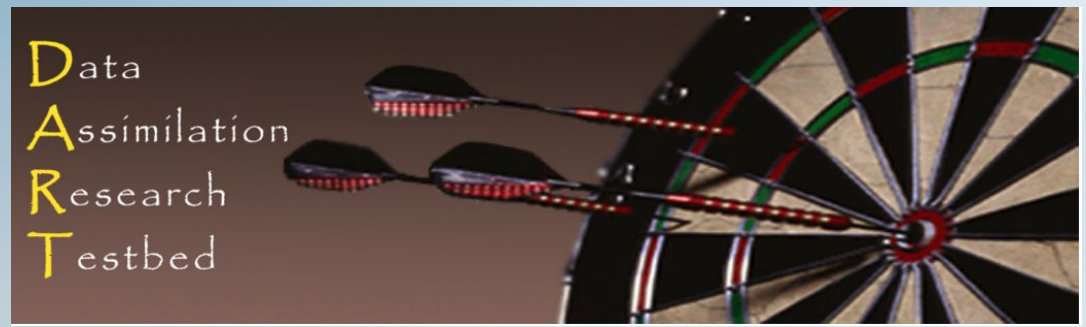


Introduction to Ensemble Data Assimilation with CLM-DART (Data Assimilation Research Testbed)

Brett Raczka, Scientist IV

Data Assimilation Research Section, NSF-National Center for Atmospheric Research (NCAR)

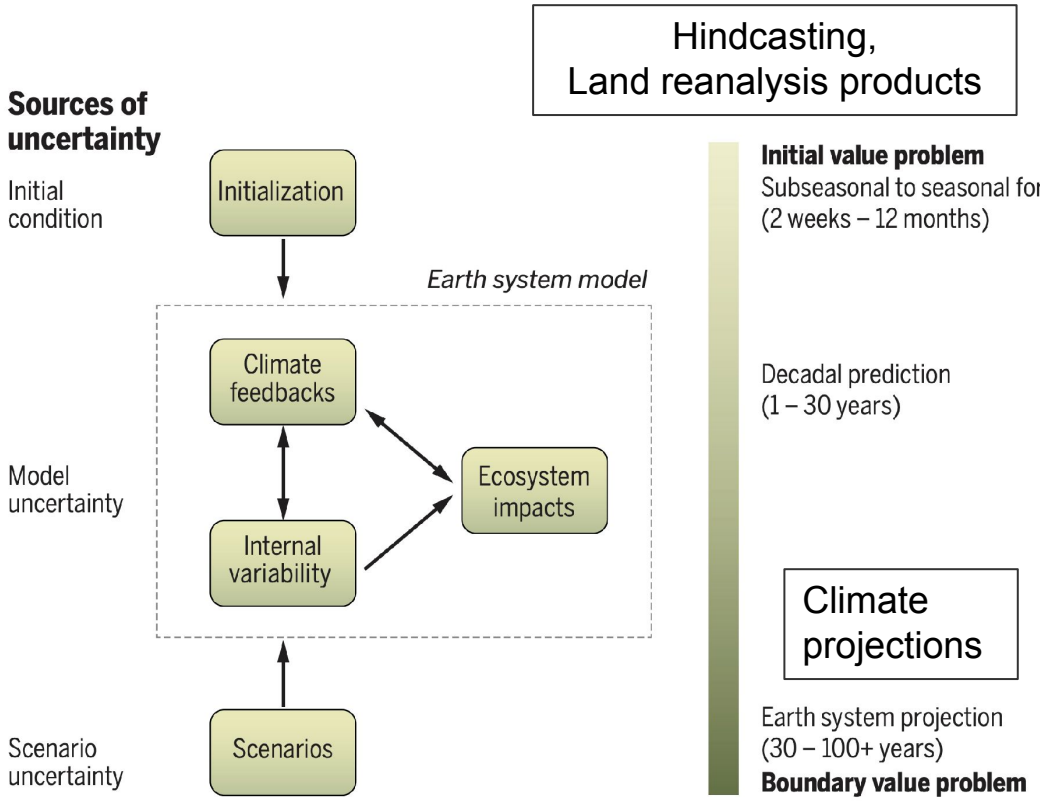
Community Land Model Tutorial

November 18, 2026

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Why use Data Assimilation (DA) for Land Models?

DA improves land model forecasts by reducing initial condition (spinup), model/parameter and boundary condition errors

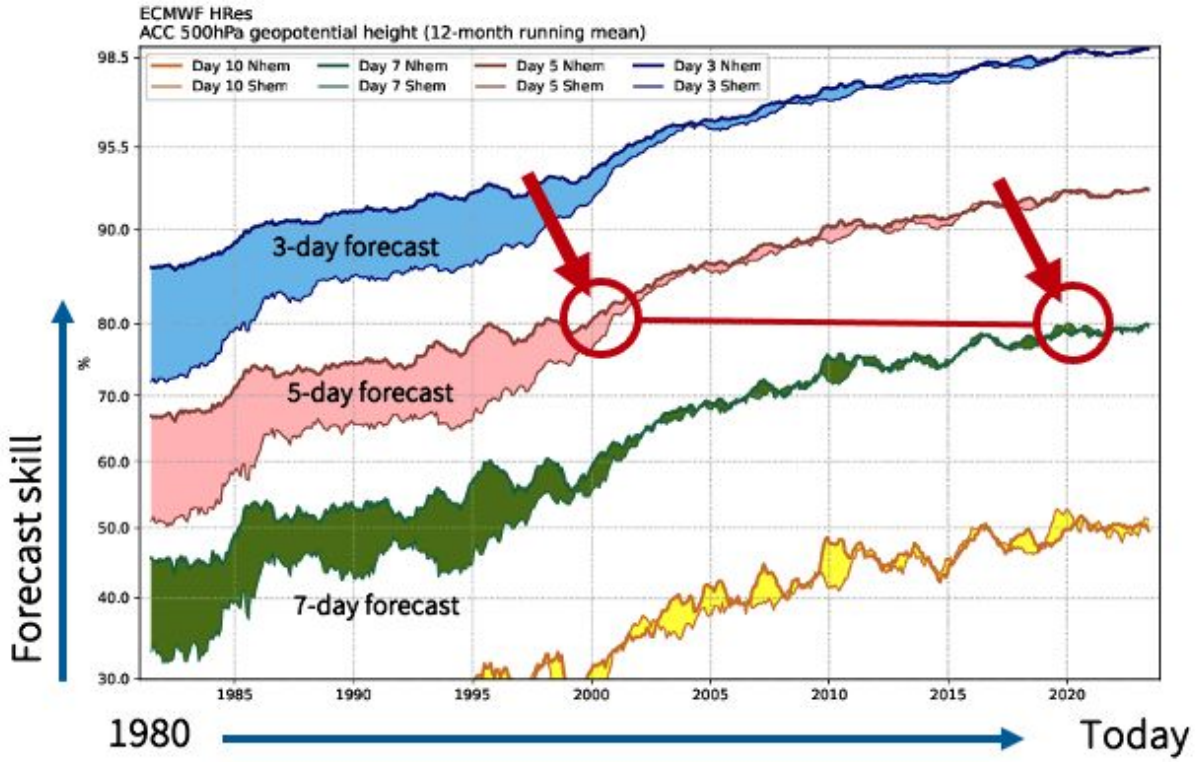


Bonan & Doney 2018

- Land Data Assimilation is mostly hindcasting, but also can be used for improving predictions (e.g. climate projections)

Similar logic as traditional Atmospheric Weather Prediction

- DA has been a key tool for improvements in atmospheric weather prediction skill
- Ingestion of observations updates atmospheric state



Forecast skill over time. Shown is the anomaly correlation of 500-hPa geopotential height forecasts (3, 5, 7, 10-day forecasts) and the verifying analysis. From ECMWF.

Fundamentals of Ensemble Data Assimilation

- Updating 'observation space' with an Ensemble Kalman Filter (EnKF)

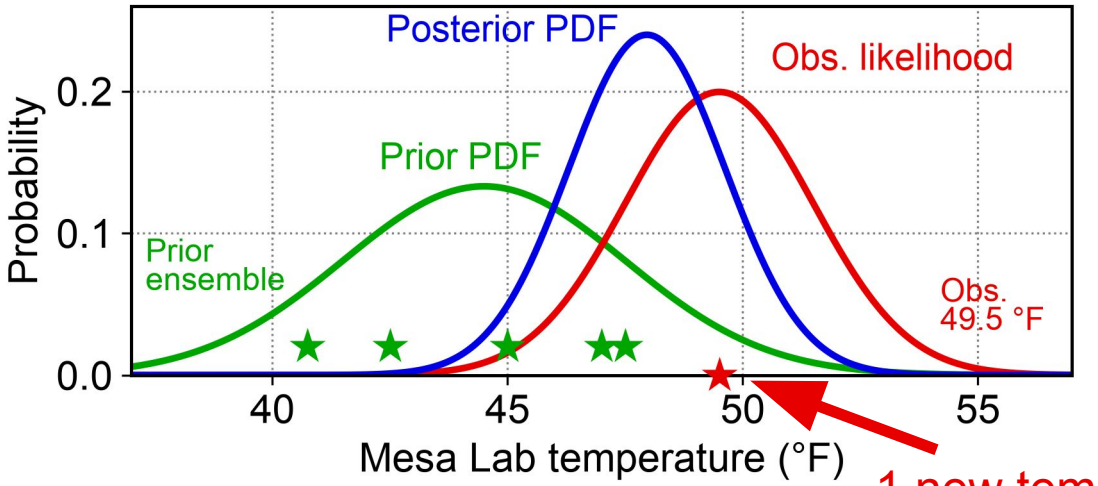
Bayes Theorem

Prior × Observation Likelihood ∝ Posterior

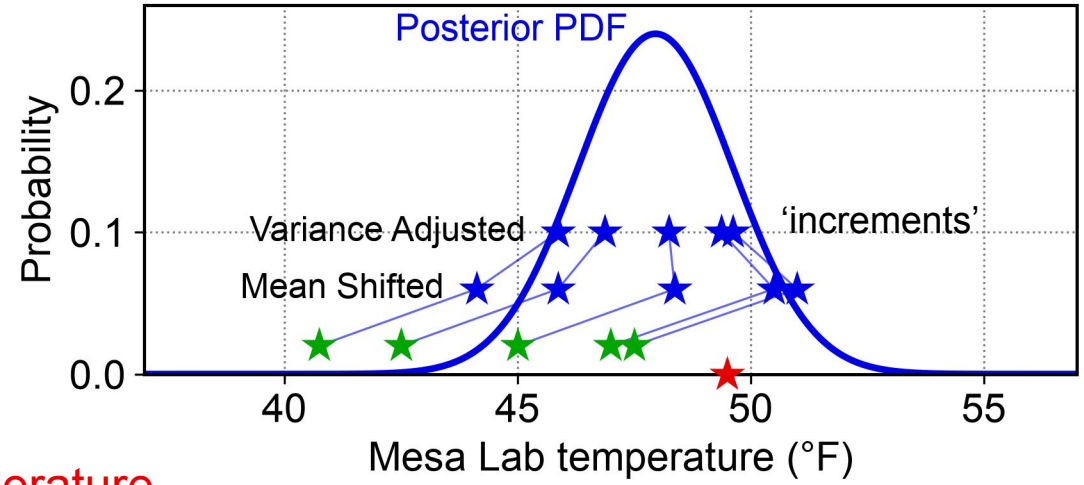
Model (CLM) generated

Land observations (e.g. temp, leaf area, soil moisture)

'Update' to CLM restart or parameter file



1 new temperature observation

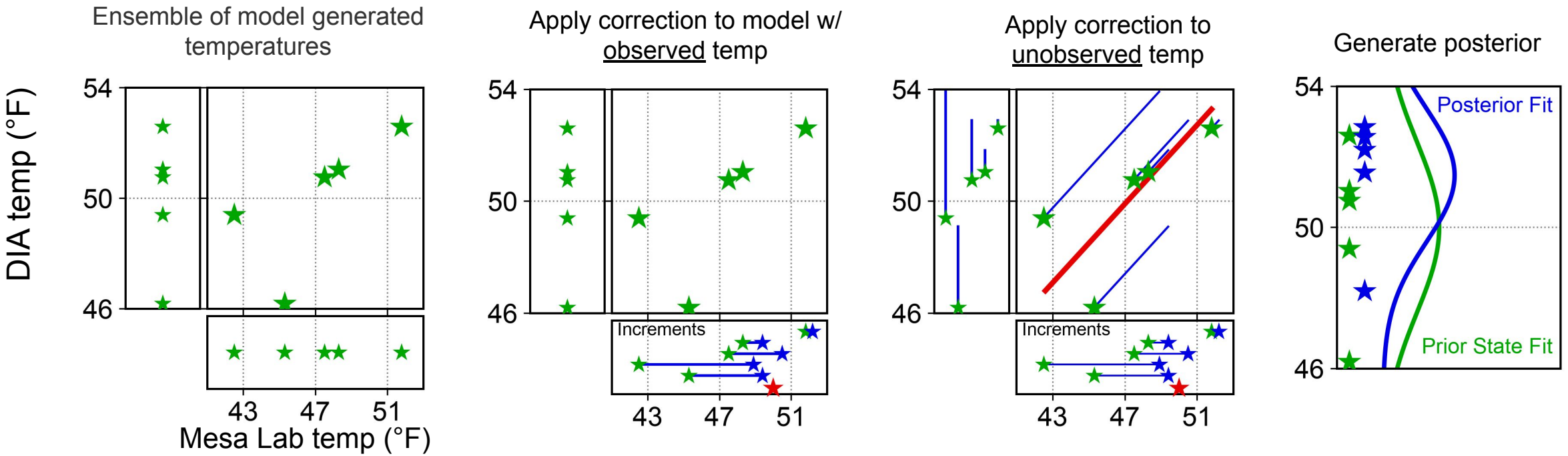


- Observations combined with prior model forecast produce an improved forecast ('posterior').
- Updates are applied to model state (e.g. temp, biomass, soil carbon) or model parameters

Fundamentals of Ensemble Data Assimilation

- Applying ‘observation space’ updates to ‘model space’

- Mid-November afternoon: we want temperature at Denver airport (DIA), but only observe at NCAR’s Mesa Lab (Boulder).



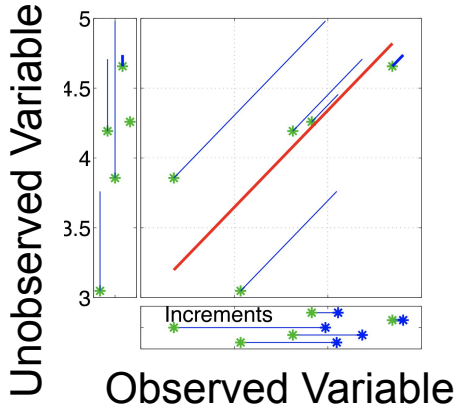
- The ensemble itself supplies the Mesa Lab–DIA correlation: members that run warm in Boulder also run warm at DIA (shared air mass). Regressing the Mesa Lab increments onto DIA updates the unobserved temperature.

- The same applies to land ! : leaf area and biomass are correlated in space (similar elevation and weather).

- This also applies *across* different land features: leaf area ~ roots, leaf area ~ soil moisture.

Important model space updates in Land DA

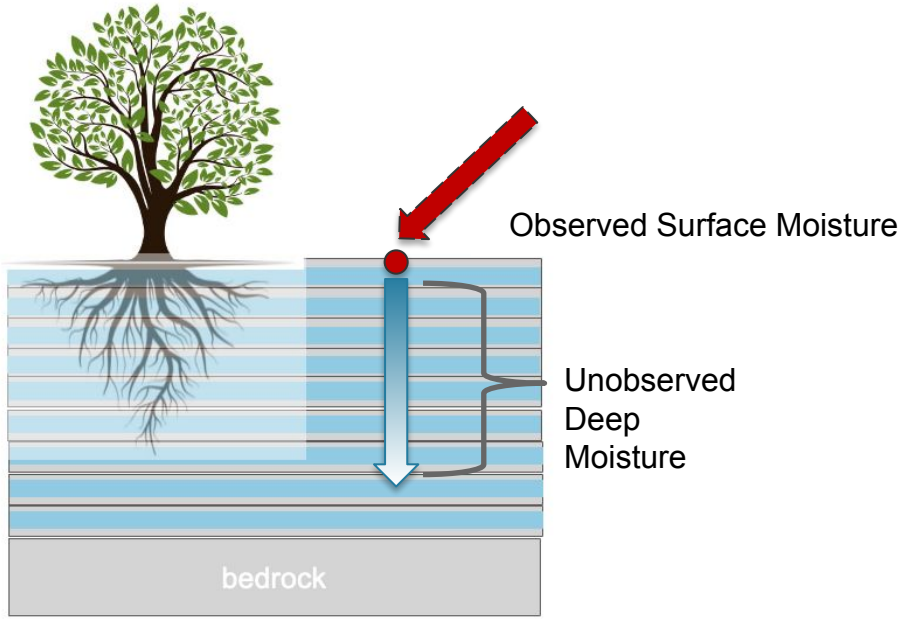
Apply model state updates



- This step is extremely useful for chronically under-observed regions of Earth System (e.g. land and ocean sub-surface)

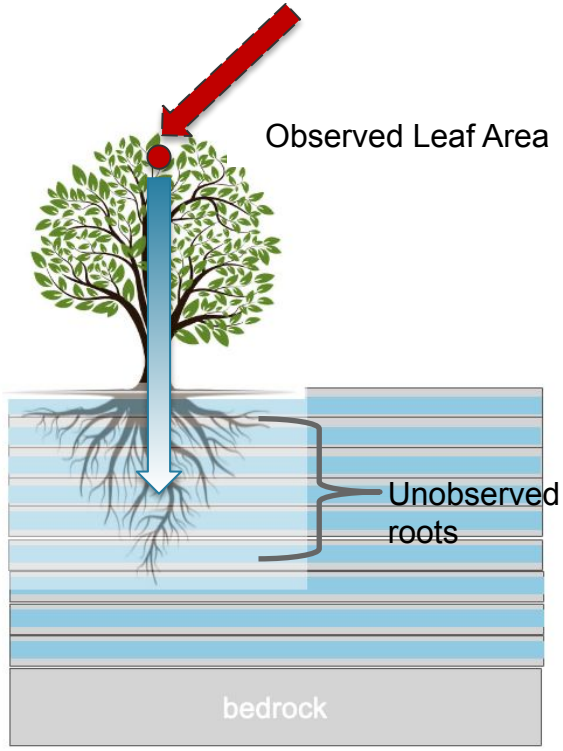
Example 1:

Surface soil moisture observations inform deeper soil moisture (Richards Equation)



Example 2:

Leaf area observations inform root carbon pool, and vertical distribution (Leaf to root allocation)



CLM-DART Overview

Data Assimilation Research Testbed (DART) and Land Modeling

A flexible set of open-source software tools to accelerate earth system research using ensemble data assimilation

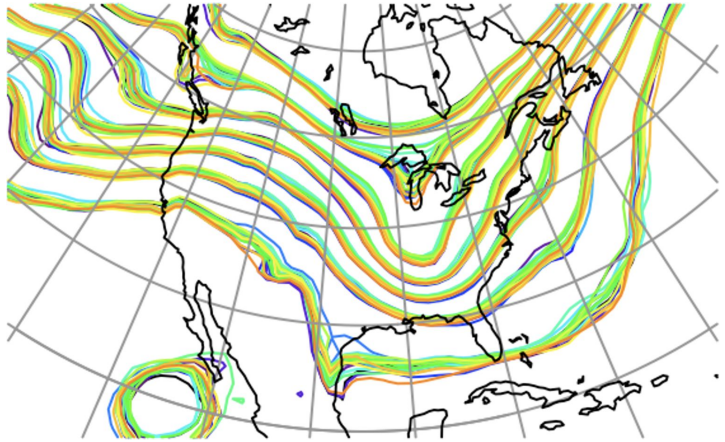
<u>DART team & community members develop</u>:

- Ensemble Data Assimilation Algorithms:
  (e.g. EnKF, RHF, Quantile Conserving Filter)
- Observation Forward Operators
- Observation Converters

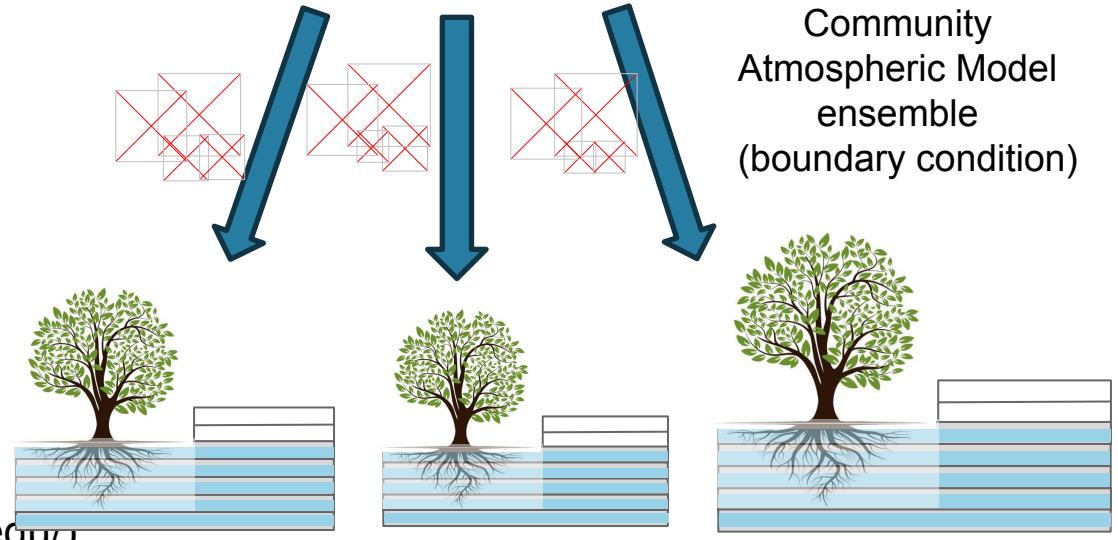
<u>Notable Land Models coupled to DART:</u>

- CESM (Community Land Model (CLM); NCAR)
- E3SM (Energy Exascale Earth System Model, (ELM); DOE)
- JULES (UK model)
- WRF-Hydro (U.S. National Water Model)
- Noah-MP
- DVM-DOS-TEM (Arctic/permafrost model)

- Educational Resource: Website, Tutorials (https://dart.ucar.edu/)



- 80 member atmospheric re-analysis (CAM) (K. Raeder 2021)



Community Atmospheric Model ensemble (boundary condition)

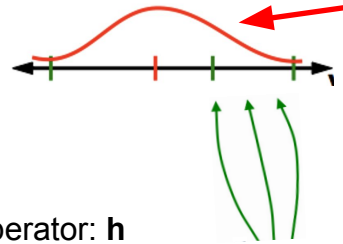
- Produces 80 member CLM ensemble for estimate of dynamic background error covariance for land DA

Standard CLM-DART Setup for ensemble (biomass example)

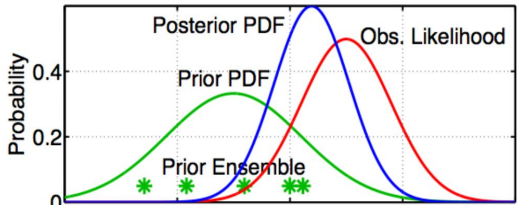
Observation Space



2 Calculate observed biomass

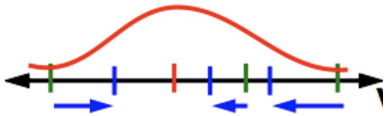


Forward operator: **h**
Observation estimate = **h**(deadstemc, livestemc, leafc, lat, lon)



State Variable

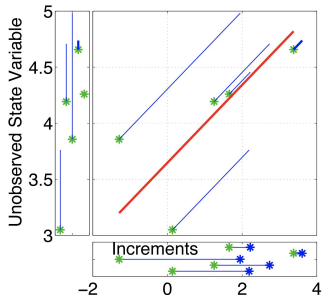
3 Update (posterior)



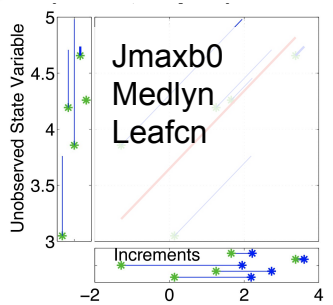
Observed Variable (biomass)

4 Apply model state updates

Deadstem(c,n)
Livestem(c,n)
Leaf(c,n)
Livecoarseroot(c,n)
Deadcoarseroot(c,n)
Fineroot(c,n)
Litter(c,n)

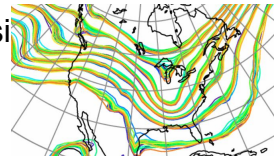


4b Apply parameter updates



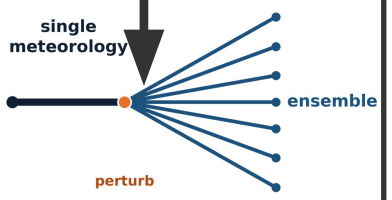
Regional Mode:

CAM Reanalysis Ensemble



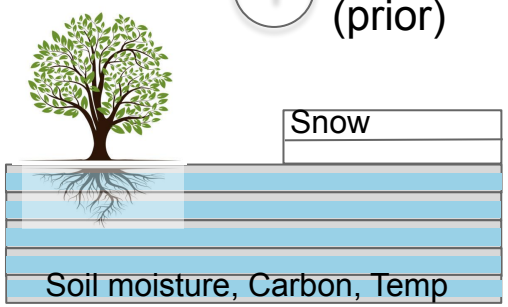
Site Mode:





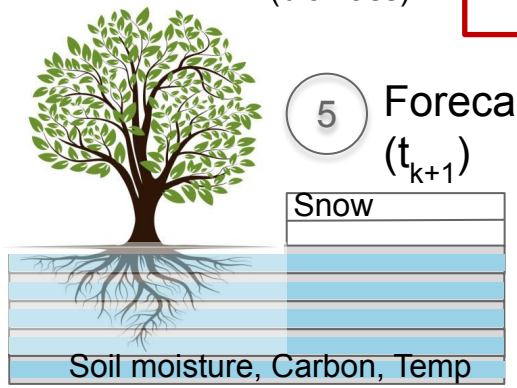
CLM History, restart files

1 Forecast (t_k) (prior)



Model Space

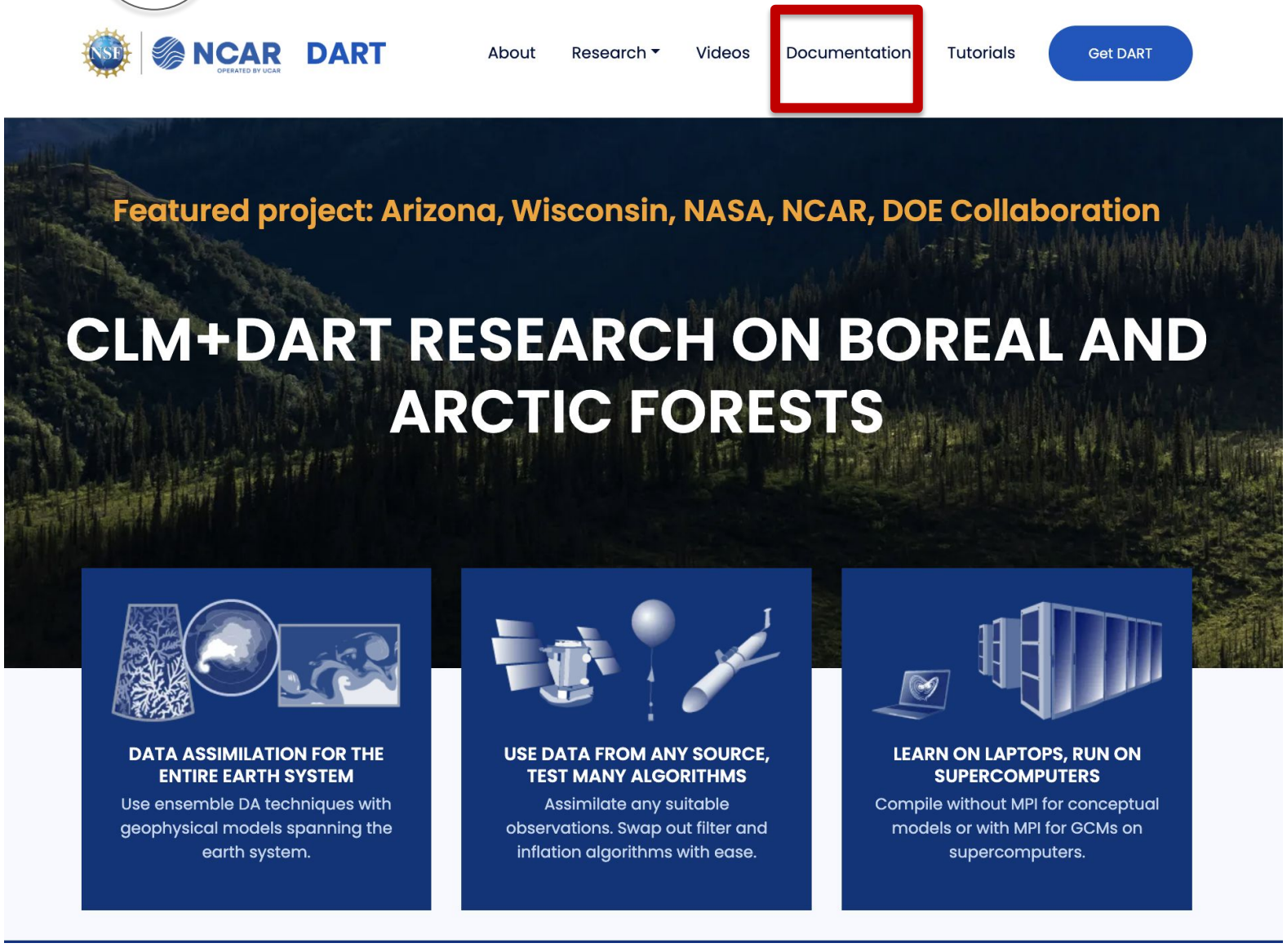
5 Forecast (t_{k+1})



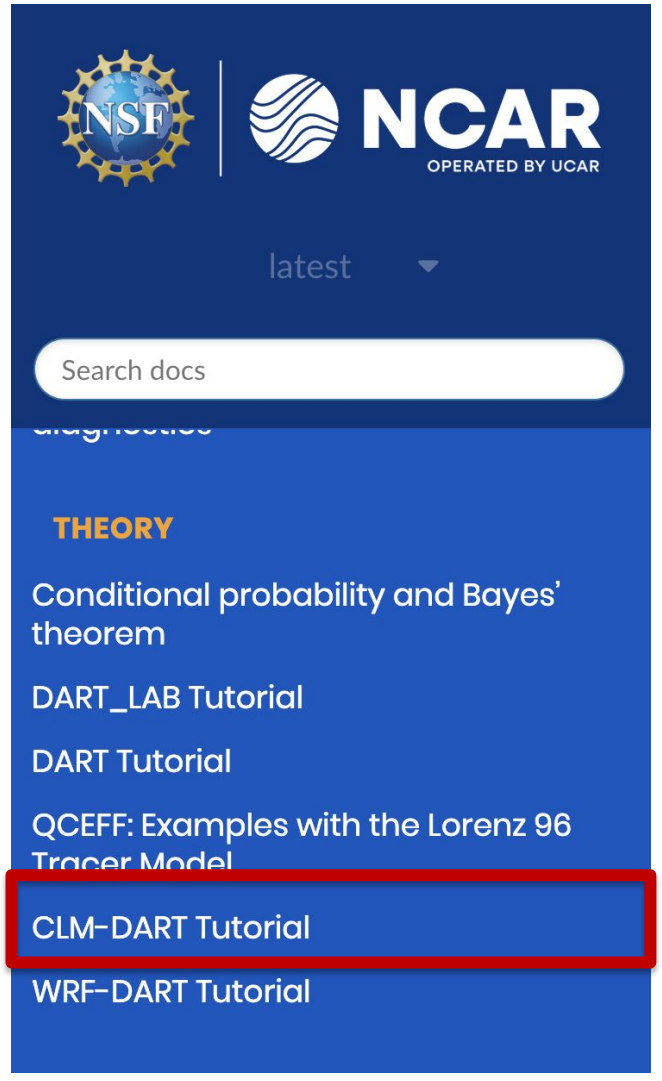
CLM-DART Tutorial Overview

Navigating to CLM-DART Tutorial

1 Website: https://dart.ucar.edu



2



CLM-DART Tutorial Overview

Get the code
Steps 1–4

1) Download CLM5 → 2) Download DART → 3) DART Scripts → 4) Compile DART

Download CLM5:
```
git clone https://github.com/ESCOMP/CTSM.git
cesm_dart
cd cesm_dart
git checkout release-cesm2.2.03
./manage_externals/checkout_externals
```

Download DART:
```
cd /glade/work/$USER/
git clone https://github.com/NCAR/DART.git
cd DART
```

Configure the assimilation
Steps 5–10

5) Atmospheric Forcing → 6) Initial Conditions → 7) Observations → 8) Forward Operator → 9) Localization → 10) Inflation

Run & diagnose
Steps 11–13

11) Complete Assimilation Setup → 12) Execute Assimilation Run → 13) Diagnose Assimilation Run

Tutorial: docs.dart.ucar.edu/en/latest/models/clm/tutorial

CLM-DART Tutorial Overview: Observations (Step 7)

Land observation converter examples:

Location: `DART/observations/obs_converters/`

Observation	DART obs type	Native Format	Location
Leaf area index MODIS	`MODIS_LEAF_AREA_INDEX`	CSV	MODIS/ `MOD15A2_to_obs`
Tower fluxes AmeriFlux FULLSET (hourly)	`TOWER_GPP_FLUX` + ER, NEE, LE, H	CSV	Ameriflux/ `fluxnetfull_to_obs`
Tower fluxes † FLUXNET2015 daily (CalLMIP)	`TOWER_NET_CARBON_FLUX` + LH, SH	netCDF	branch (not in main): `calLMIP_assim`
Soil moisture SMAP L2	`SOIL_MOISTURE`	HDF5	NSIDC/ `SMAP_L2_to_obs`
Soil moisture LPRM	`LPRM_SOIL_MOISTURE`	netCDF	NASA_Earthdata/ `LPRM_L3_to_obs`
SIF GOME-2 / SCIAMACHY	`HARMONIZED_SIF`	netCDF	SIF/ `SIF_to_obs_netcdf`
Snow cover fraction MODIS	`MODIS_SNOWCOVER_FRAC`	ASCII	snow/ `snow_to_obs`
Soil organic carbon † SoilGrids (0–1 m stock)	`SOIL_CARBON_1m`	ASCII	branch (not in main): `calLMIP_assim`

† calLMIP_assim branch, not yet in main DART.
Full list: docs.dart.ucar.edu/en/latest/observations/obs_converters/README.html

`DART obs_seq.out` format

ASCII formatted

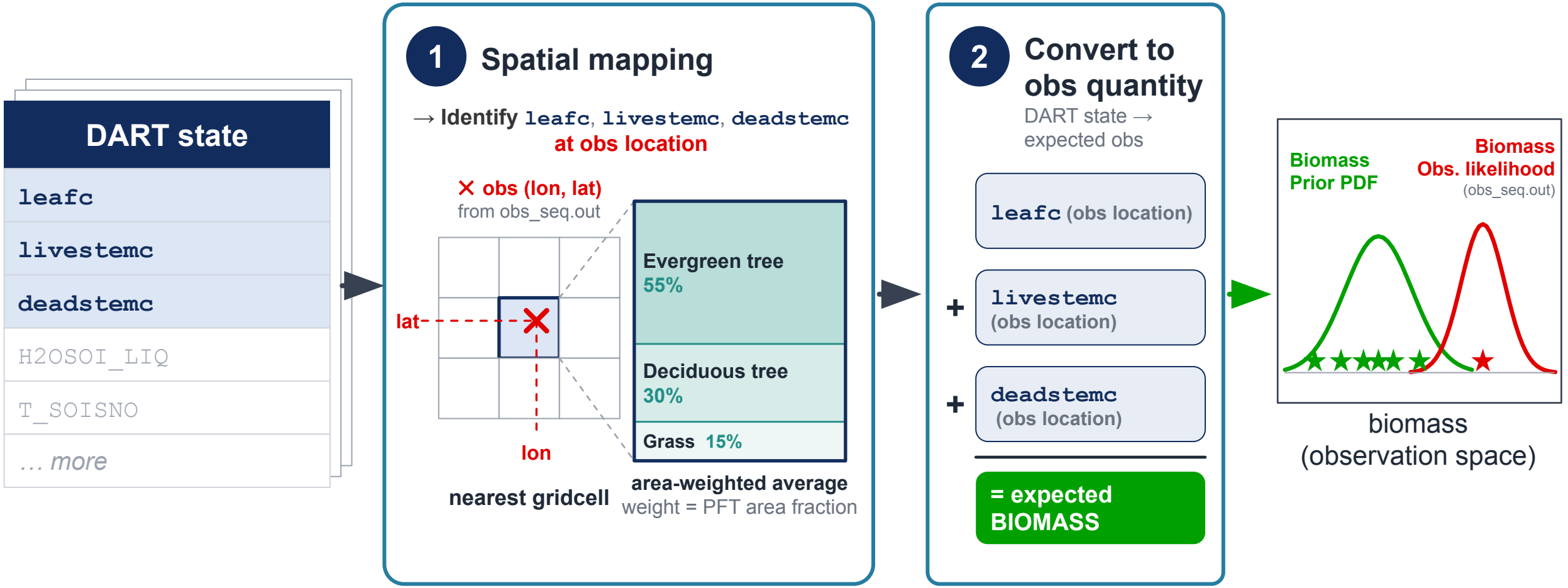
```
obs_sequence
obs_type_definitions
   5
   1 SOIL_TEMPERATURE
  22 MODIS_SNOWCOVER_FRAC
  23 MODIS_LEAF_AREA_INDEX
  28 BIOMASS
  42 HARMONIZED_SIF
 num_copies: 1   num_qc: 1
 num_obs: 5   max_num_obs: 5
observations
Quality Control
 first: 1   last: 5
 OBS   5
   5.0007441551129297
   0.0000000000000000
   4   -1   -1
obdef
loc3d
   4.4222   0.4211   -999.0
-2
kind
   23
   0   146827
   4.0000000000000008E-02
```

- **Obs types in this file** (number ↔ name)
- **Copies per obs + QC fields**
- **Number of observations**
- **Copy name + QC name**
- ▼ **Showing OBS 5 only**
- **Observed value**
- **QC flag**
- **Observation Order**
- **3-D location:**
- **lon, lat (rad), vertical, type**
- **Obs type 23** = MODIS LAI
- **Seconds, days** (2003-01-01)
- **Observation error variance**

CLM-DART Tutorial Overview: Forward Operator (Step 8)

The 'forward operator' uses the DART state to calculate the expected observation (prior):

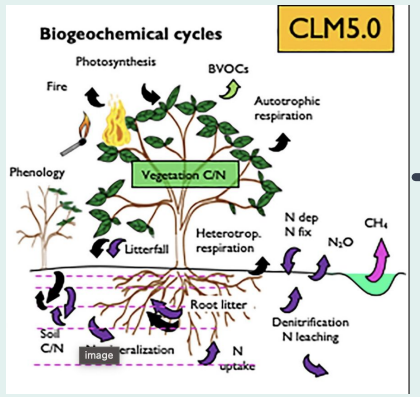
Example: `Aboveground BIOMASS observation` = leaf carbon + live stem carbon + dead stem carbon at the observation location



CLM-DART Tutorial Overview: Assimilation (Step 12)

CLM

START



restart.nc
leafc
frootc
H2OSOI_LIQ
cpool
T_SOISNO
deadstemc
… many more

END

updated restart.nc
leafc*
frootc
H2OSOI_LIQ*
cpool
T_SOISNO*
deadstemc*
… many more

UPDATE variables copied back for the next CLM forecast

*** = posterior values**

DART

Apply variable localization

DART state
leafc
H2OSOI_LIQ
T_SOISNO
deadstemc

forecast.nc

Apply Prior Inflation



preassim.nc

DART 'filter' stages

analysis.nc

output.nc

Clamp state values

Apply Posterior Inflation



postassim.nc

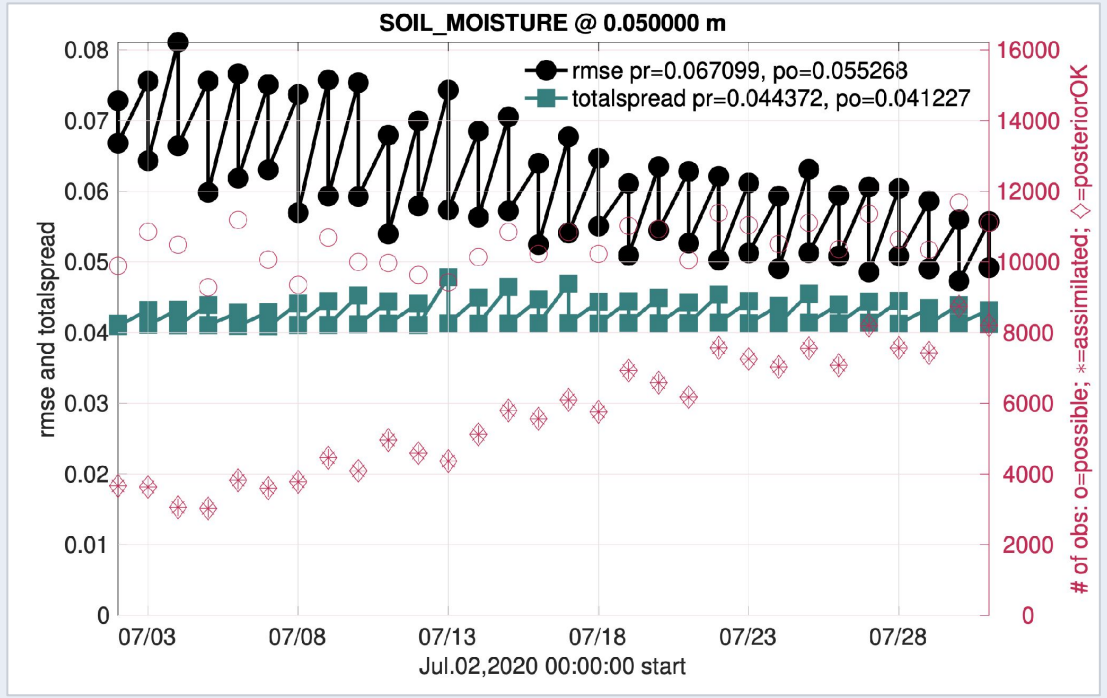
Assimilation Updates State



× = observation location

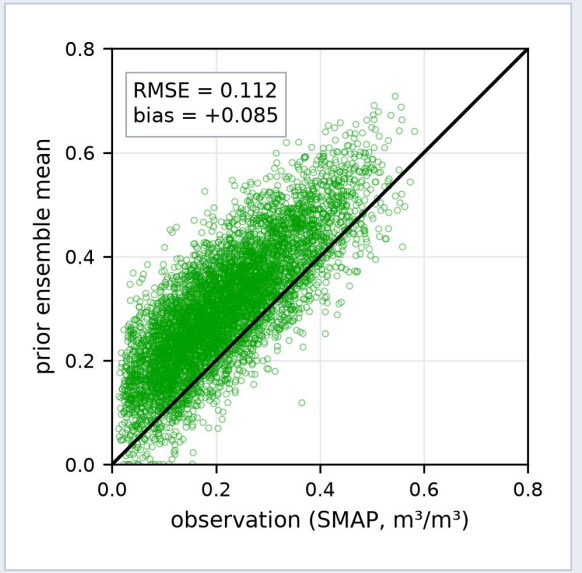
CLM-DART Tutorial Overview: Diagnostics (Step 13)

Example 1 RMSE evolution `plot_rmse_xxx_evolution.m`

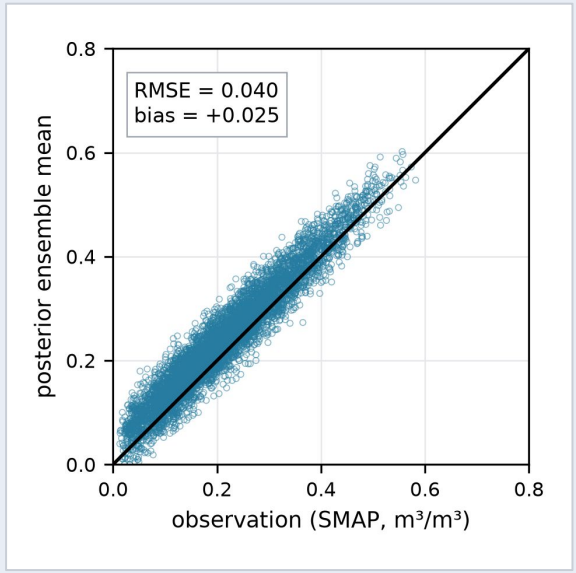


● RMSE ↓ ■ total spread ≈ RMSE ◇ assimilated obs ↑

Example 2 1:1 plot `link_obs.m` *ensemble mean vs. obs*



Prior (forecast)
wet bias, broad scatter



Posterior (analysis)
tight along the 1:1 line

Diagnostic	MATLAB in DART download	pyDARTdiags Python package
RMSE / spread over time	`plot_rmse_xxx_evolution.m` ← `obs_diag`	`mp.plot_evolution(obs_seq, …)`
1:1 ens. mean vs. obs	`link_obs.m` ← `obs_seq_to_netcdf`	
Profiles, rank histogram	`plot_profile.m`, `plot_rank_histogram.m`	`mp.plot_profile`, `mp.plot_rank_histogram`

MATLAB scripts come with DART

```
git clone https://github.com/NCAR/DART.git
cd DART/diagnostics/matlab
```

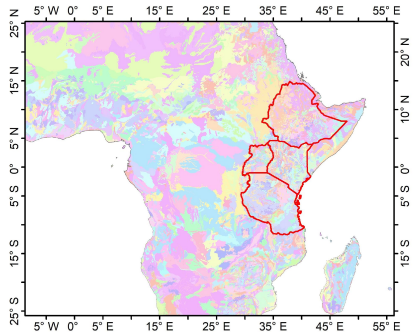
pyDARTdiags separate install

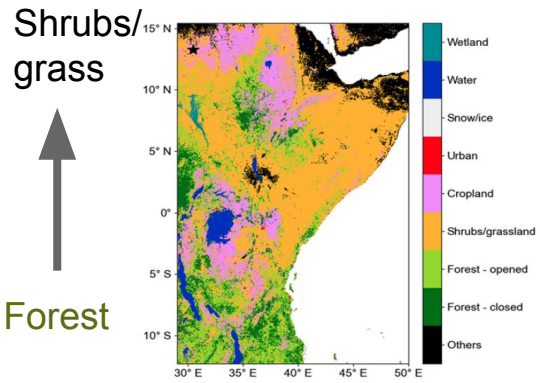
```
pip install pydartdiags
```

github.com/NCAR/pyDARTdiags

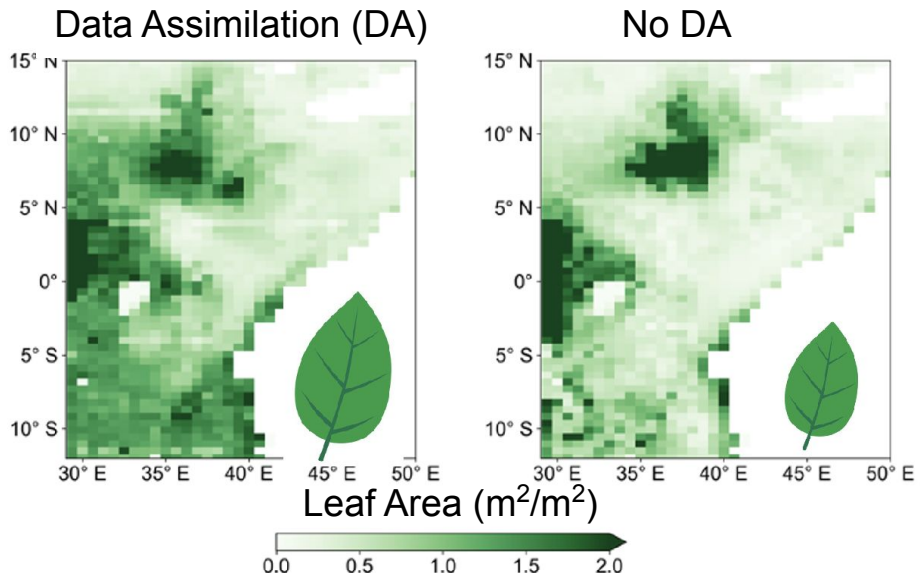
CLM-DART Applications:
State Estimation

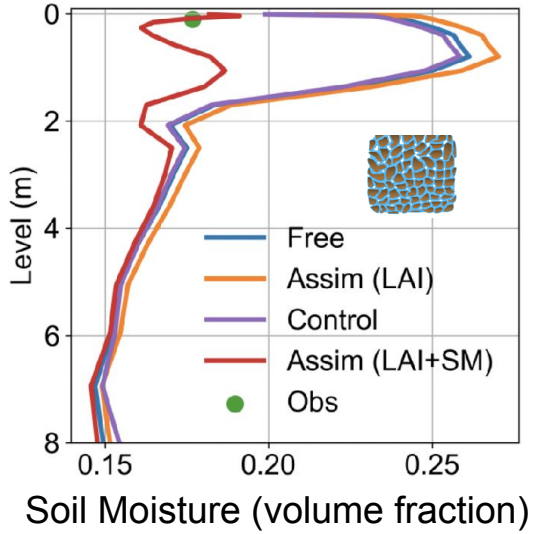
Climate: Carbon Monitoring Systems, Forest Restoration (Nature Based Solutions)





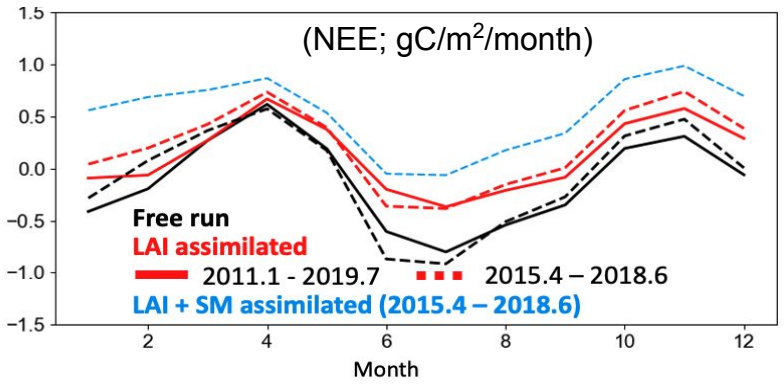
- Informing land restoration practices for remote & vulnerable ecosystems with MODIS leaf area and SMAP soil moisture observations



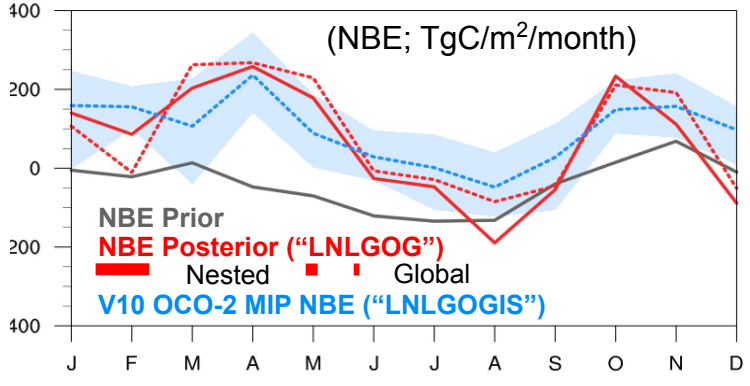


- Leaf area and soil moisture observations bring the 'bottom up' and 'top down' estimate of carbon exchange into better agreement

E. Africa 'Bottom Up' Carbon Exchange



Africa 'Top Down' Carbon Exchange



Ying Sun
Jiameng Lai
Zhenqi Luo

Cornell University

Hydrology: Snow and water limitation on forest productivity

- Snow DA requires repartitioning algorithm to account for variable snow layers across ensemble members, and for zero spread (no snow condition)

- DA increases high elevation snow, enhances SM and increases carbon uptake



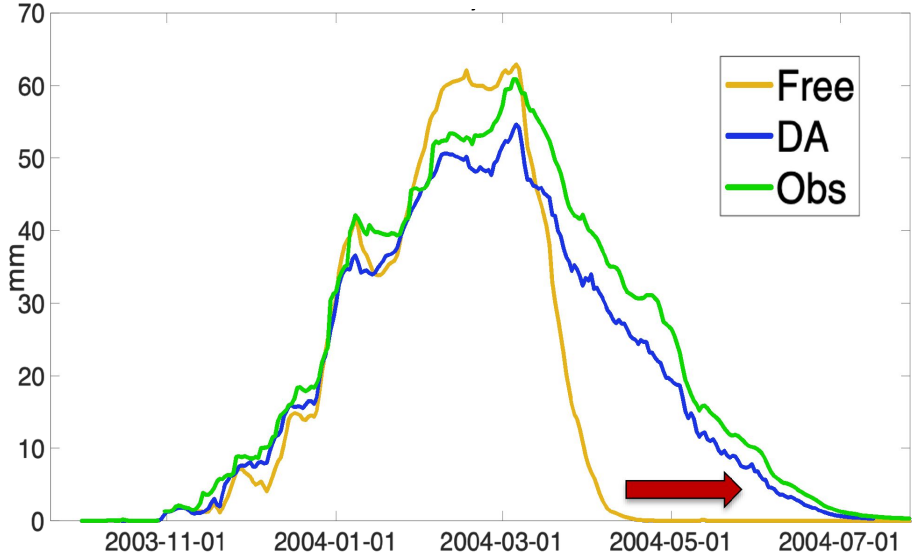


THE UNIVERSITY OF UTAH

Xueli Huo et al., (under revision, JGR-Biogeosciences)

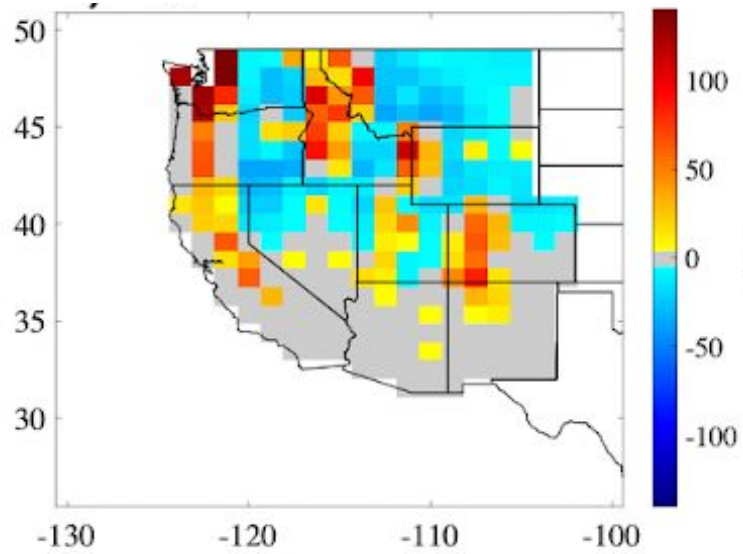
Western United States simulated snowpack (CLM, CLM-DART)

Seasonal snow water equivalent



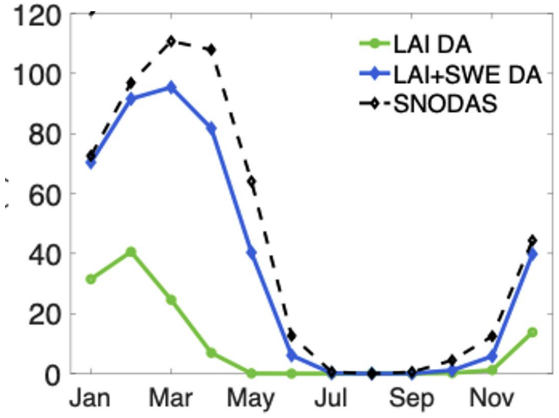
- Early spring melt-off corrected

DA - Free Run

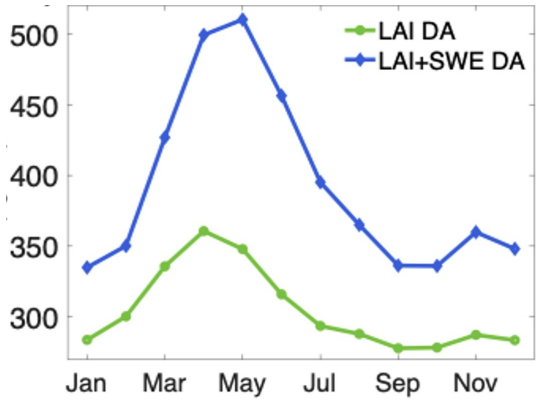


- Snow redistributed from low to high elevations

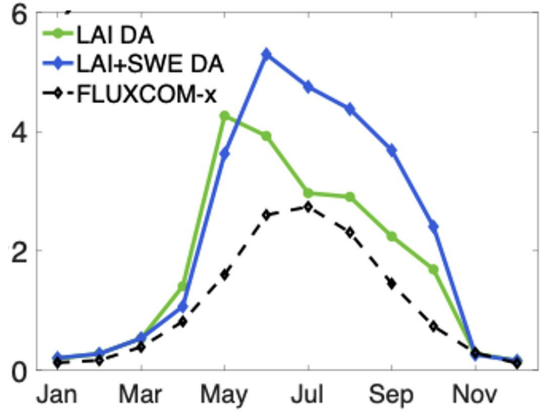
Snow Water Equivalent (mm)



RZ Soil Moisture (kg/m^2)



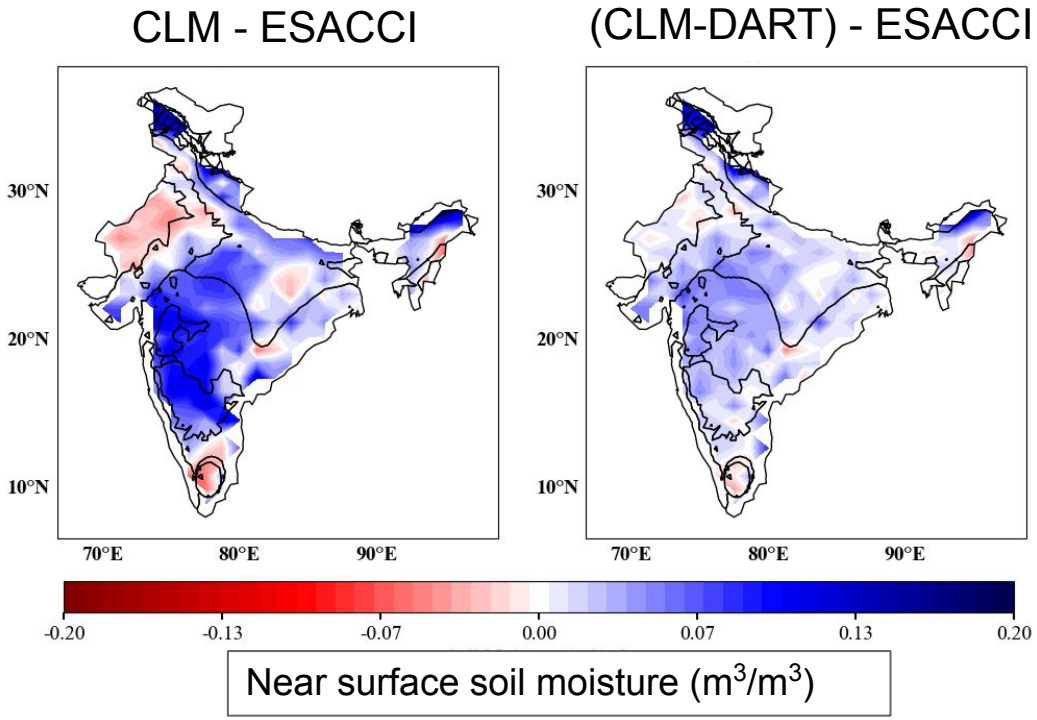
Gross Primary Productivity ($gC/m^2/day$)



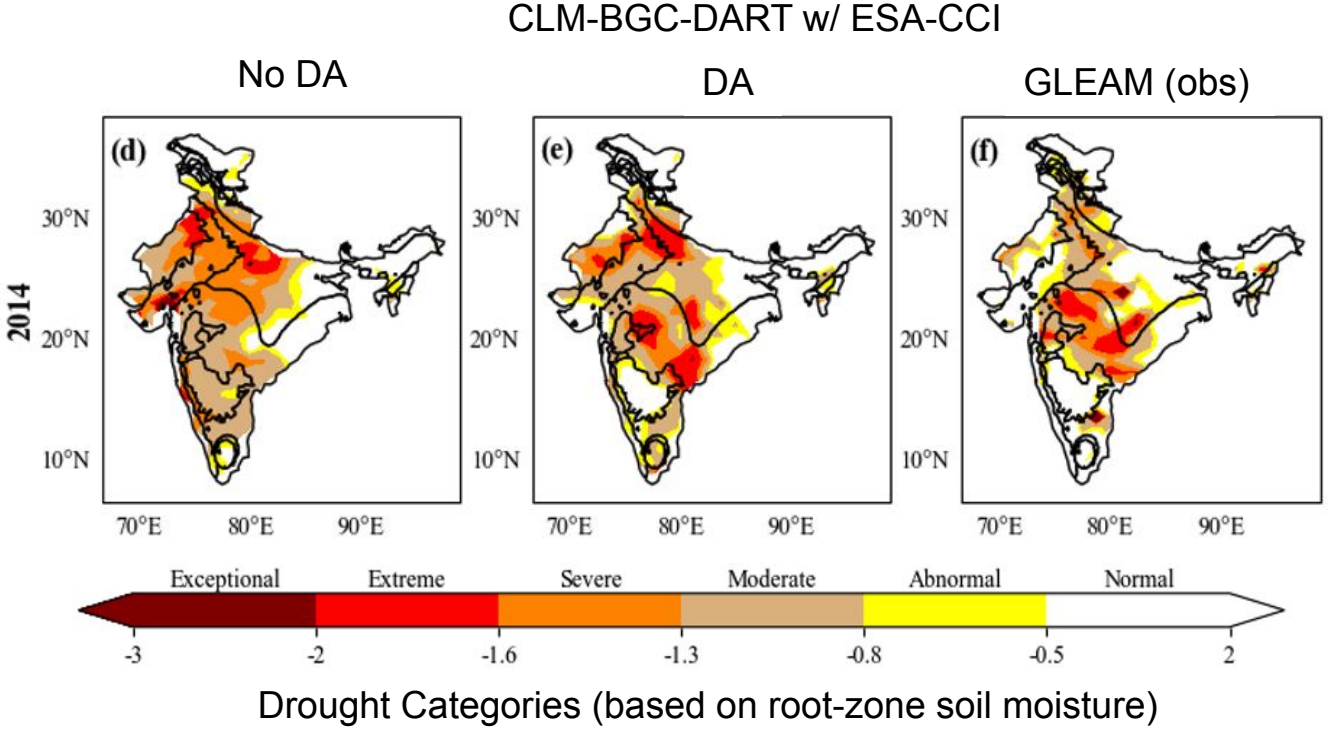
Hydrology: Drought Monitoring

Assimilating **ESA-CCI satellite soil moisture** into CLM with DART gives an improved estimate of **drought severity and extent** across India.

① DA removes most of CLM's wet bias



② With DA, 2014 drought pattern moves closer to GLEAM (benchmark)

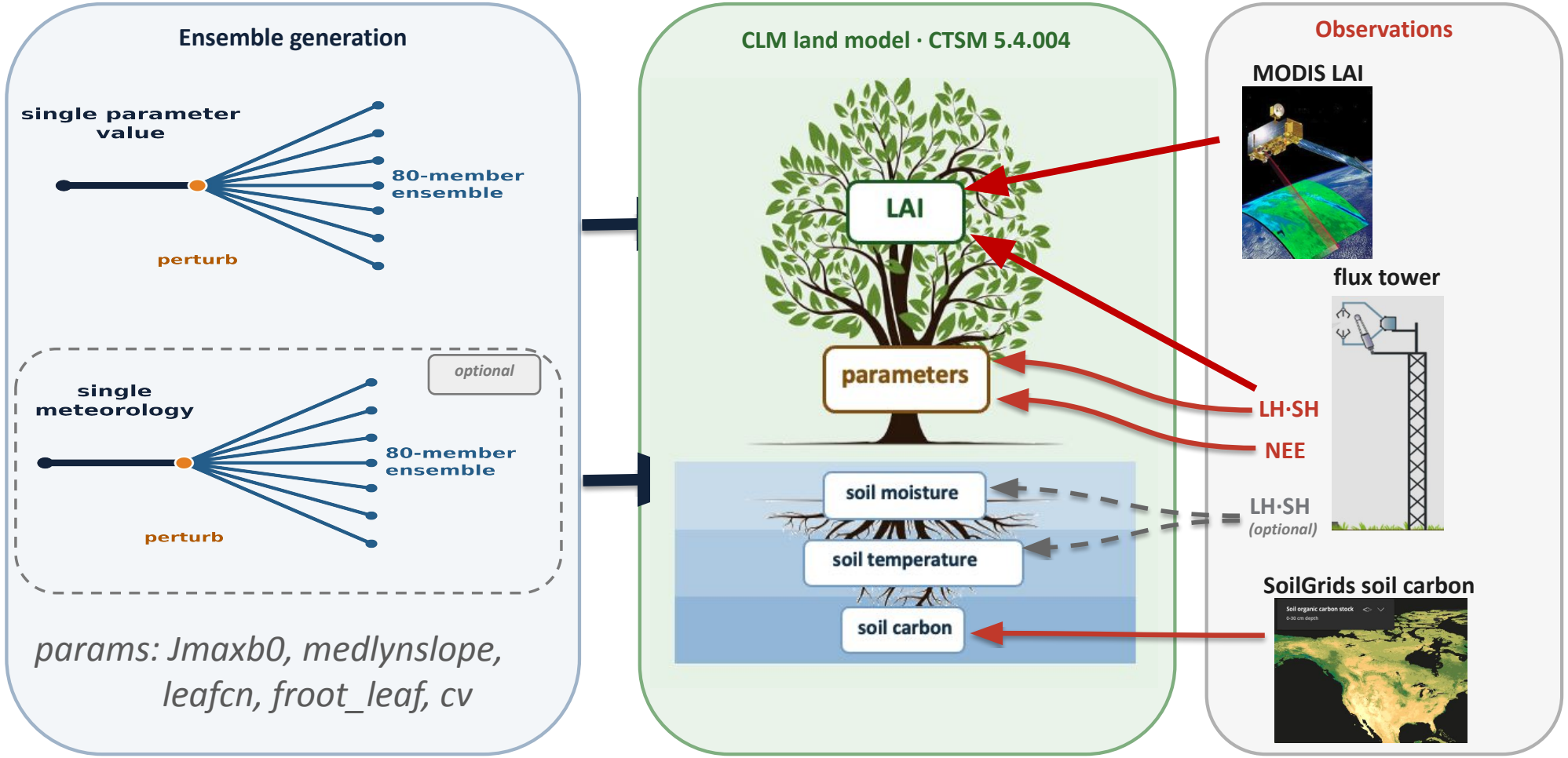


Collaboration with Chiru Devavat (HydroX lab, IIT Delhi)

CLM-DART Applications:
State + Parameter Estimation

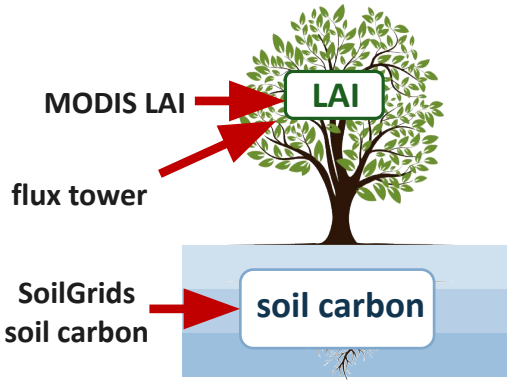
Addressing structural/parametric error: Joint parameter & state estimation

- (CalLMIP): quantify uncertainty in carbon cycle projections across models using parameter estimation
- Leverage flux tower observations (carbon, energy) to improve models

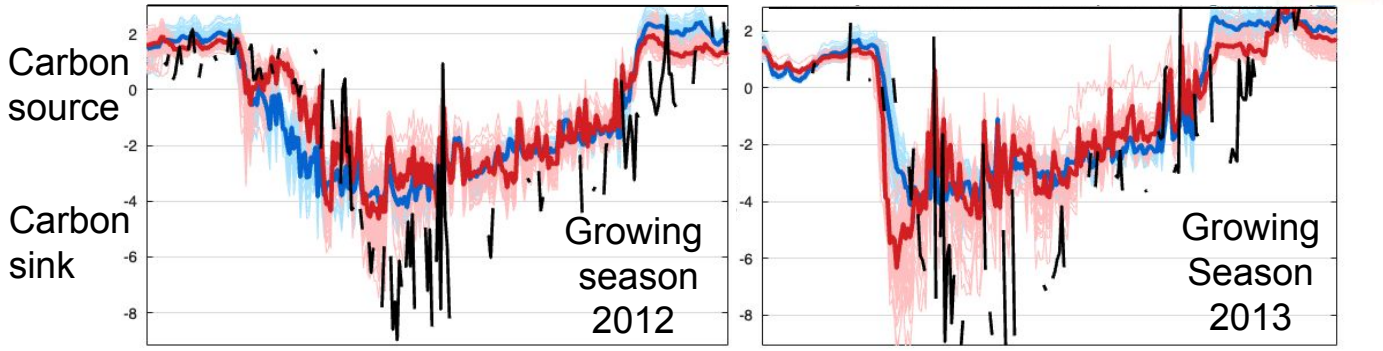


Single-point DK-Sor (beech forest) · 80-member CLM–DART · 2-day assimilation frequency

Parameter estimation improves seasonal carbon uptake

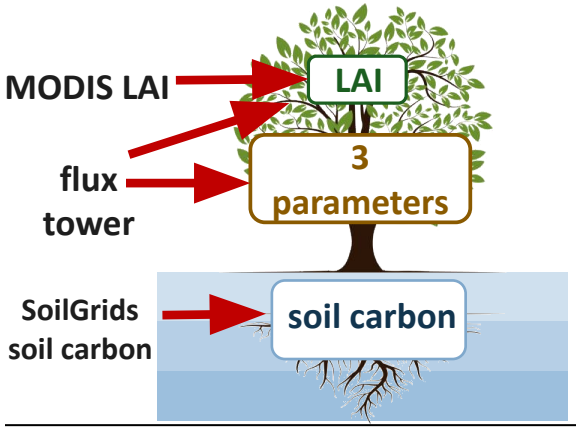


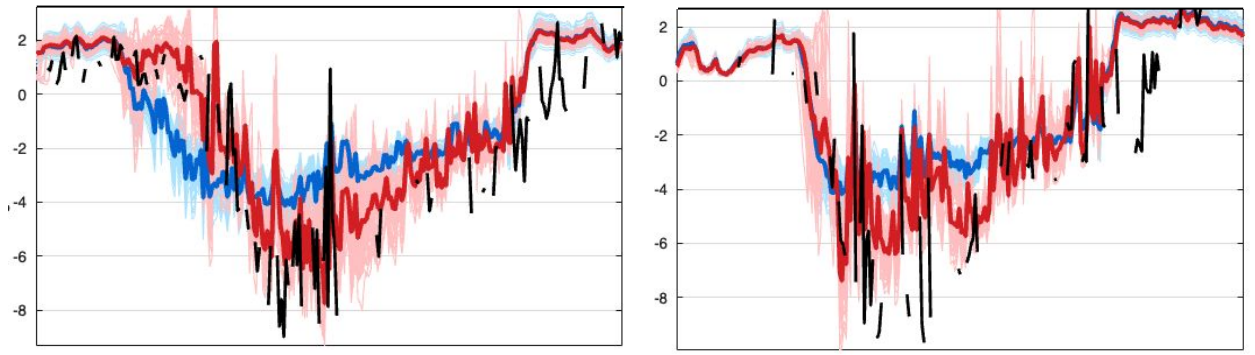
Net carbon exchange (NEE, gC/m^2/day) Free: DA forecast:



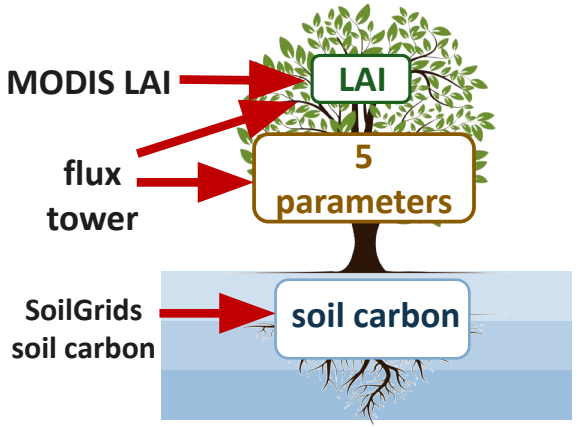
RMSE forecast (prior)

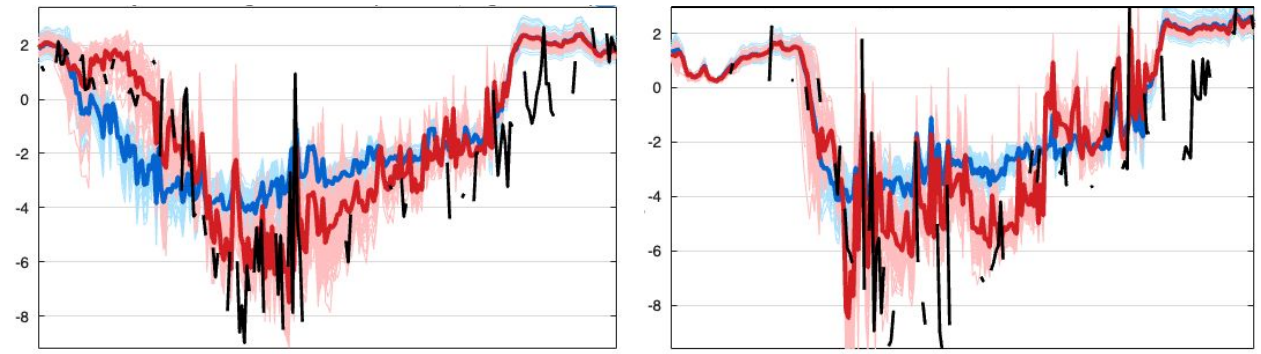
LAI (m^2/m^2)	1.43	0.60
NEE (gC/m^2/d)	2.40	2.49
LH (W/m^2)	21.5	21.5
SH (W/m^2)	43.3	45.3





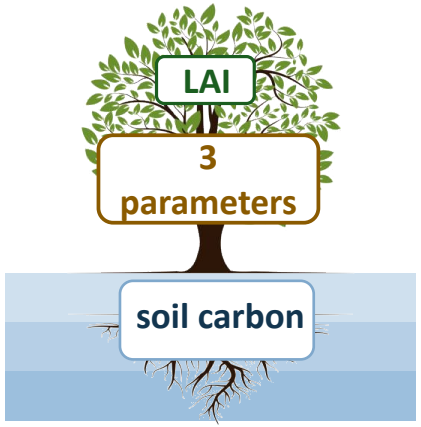
LAI (m^2/m^2)	1.43	0.69
NEE (gC/m^2/d)	2.40	1.70
LH (W/m^2)	21.5	19.7
SH (W/m^2)	43.3	43.3





LAI (m^2/m^2)	1.43	0.77
NEE (gC/m^2/d)	2.40	1.70
LH (W/m^2)	21.5	19.0
SH (W/m^2)	43.3	42.7

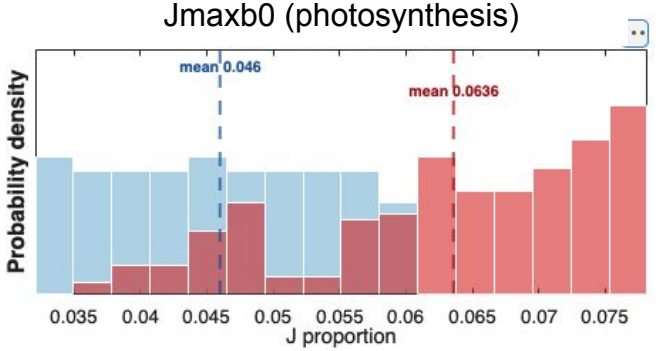
Improved parameter distributions can lead to improved forecasts



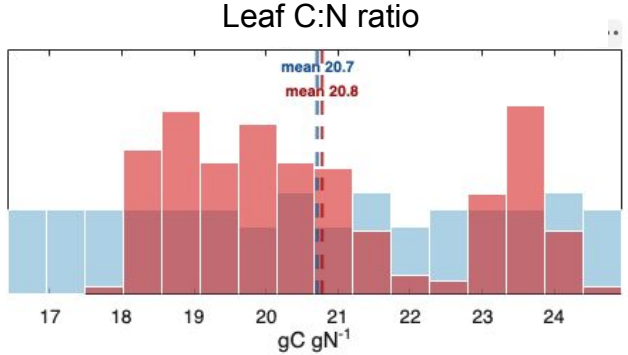
- Flux data provides significant constraint and shifts to photosynthesis and stomatal conductance parameters

Prior: Posterior:

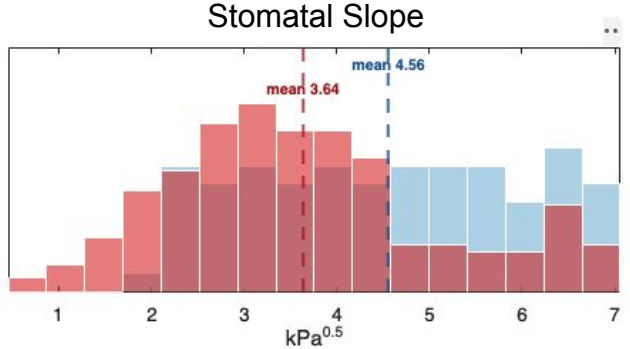
Jmaxb0 (photosynthesis)

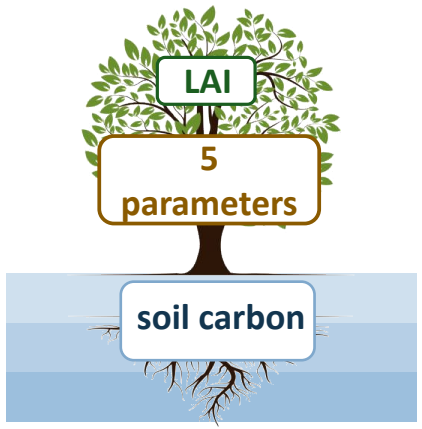


Leaf C:N ratio

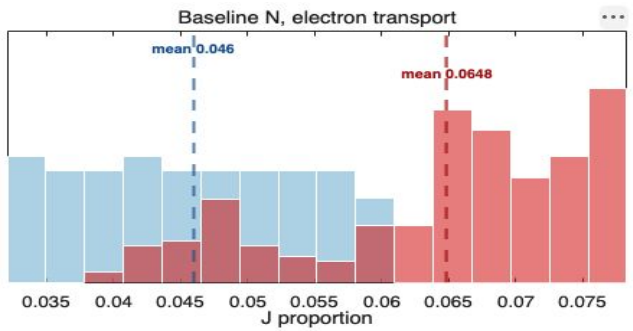


Stomatal Slope

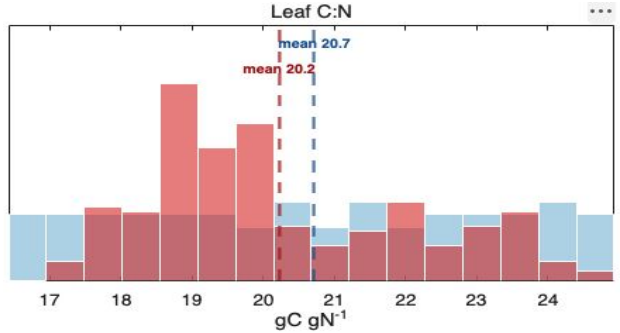




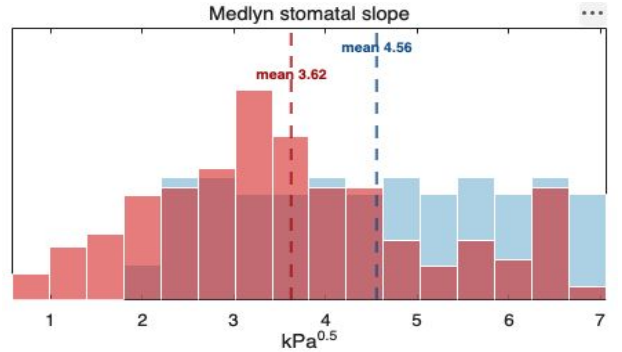
Baseline N, electron transport



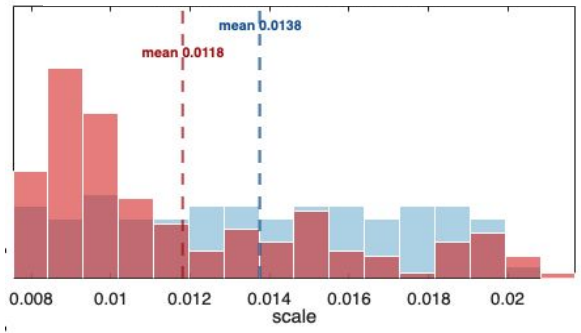
Leaf C:N



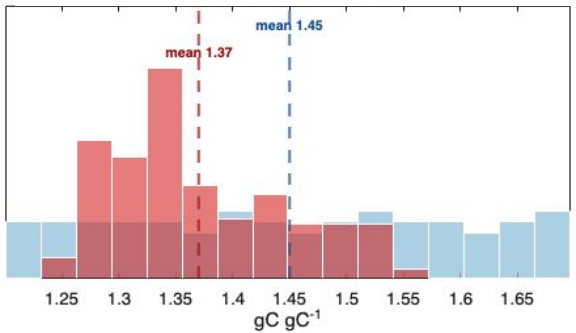
Medlyn stomatal slope



Leaf/Canopy Transfer Coefficient



Root:Leaf allocation



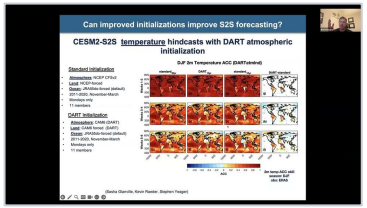
Recorded Talks: CLM-DART Applications

Questions related to Land DA and CLM-DART: bmraczka@ucar.edu or dart@ucar.edu

AIMES Land Data Assimilation Workshop *AIMES YouTube channel*

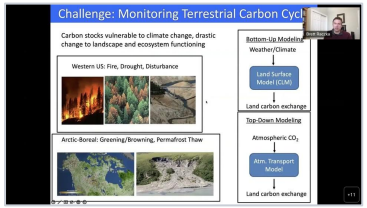
Recording coming soon

2026 · 6th workshop
Oct 2026
Challenges in Representing Land-Atmosphere Processes using DART
Recording pending



2025 · 5th workshop
Oct 2025 · 10 min
Improving Land-Atmosphere Forecasts using DART
▸ Watch on YouTube

ISDA Online *RIKEN Data Assimilation channel*



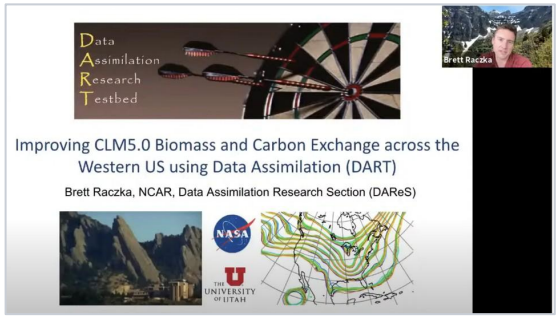
Dec 2022
20 min
Improving the Representation of Land Surface Processes using DART
▸ Watch on YouTube

NSF NCAR recordings *NCAR CGD YouTube channel*

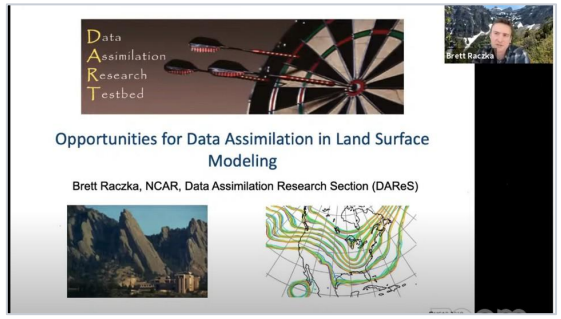
CESM Workshop *working-group session recordings*



2022 · LMWG
Jun 2022 · talk starts at 3:35:40
Applying DART to improve Earth System Carbon, Water and Energy Cycling
▸ Watch on YouTube

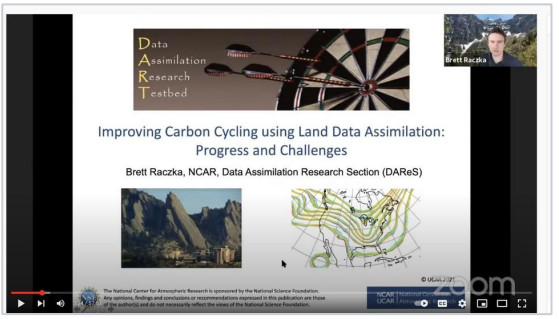


2021 · BGCWG
Jun 2021 · talk starts at 1:51:40
Improving CLM5.0 Biomass and Carbon Exchange across the Western US
▸ Watch on YouTube



2021 · LMWG
Jun 2021 · talk starts at 2:34:20
Opportunities for Data Assimilation in Land Surface Modeling
▸ Watch on YouTube

CGD Seminar Series



Nov 2021
58 min
Improving Carbon Cycling using Land Data Assimilation
▸ Watch on YouTube

Ways to collaborate with NSF NCAR

FACULTY & VISITING RESEARCHERS

NCAR Faculty Fellowship

1–3 month visit with an NCAR host in any lab

- U.S. faculty at any career stage
- Travel + $7,500/mo living allowance
- Annual call (opens Aug, due Oct)
- Hosts include **data assimilation** staff

CISL Visitor Program

2 weeks–3 months with a CISL collaborator

- Any career stage, rolling applications
- Focus: **data assimilation**, HPC & ML
- Living allowance if funds allow

Faculty Innovator Program

2-year funded research team with NCAR scientists

- Up to $125K/yr for summer salary & students

AI Expert-in-Residence

NEW

New program

- Launches November 2026

POSTDOCS

ASP Postdoctoral Fellowship

2 years of independent research at NCAR

- ~$83K/yr + travel · all disciplines

Climate & Global Change Postdoc

2-year UCAR climate postdoc

- Host anywhere in the U.S.
- Topics include **data assimilation**

STUDENTS

SIParCS

11-week paid summer internship in computing

- $23–28/hr, housing & travel covered
- Apply by Jan 20, 2027

Details: edec.ucar.edu · cisl.ucar.edu/outreach