Improving Carbon Cycling using Land Data Assimilation: Progress and Challenges

Brett Raczka, NCAR, Data Assimilation Research Section (DAReS)
Overview

• Theory/Methods of EnKF Data Assimilation, Data Assimilation Research Testbed (DART)

• Application of Data Assimilation to Western US Carbon Cycling

• Future Directions: expanding satellite observations of land surface properties

Stavros et al., (2017)
Motivation for DA in Earth System Models

Bonan & Doney 2018

**Sources of uncertainty**
- Initial condition
  - Initialization
- Model uncertainty
  - Climate feedbacks
  - Ecosystem impacts
  - Internal variability
- Scenario uncertainty
  - Scenarios

**Earth system model**

**Initial value problem**
- Subseasonal to seasonal forecast (2 weeks – 12 months)

**Boundary condition uncertainty**

**Decadal prediction**
- (1 – 30 years)

**Earth system projection**
- (30 – 100+ years)

**Boundary value problem**

Simulated Biomass in Western US
Duarte et al., (in revision)

- Elevation:
  - $< 2200$ m
  - $> 2200$ m

AGB (Pg C)

Boundary Condition Uncertainty
Model Uncertainty
Motivation for DA in Earth System Models

Dietze et al., 2018
Earth System DART applications

Atmosphere: CO w/ CAM-Chem
Gaubert et al., (2020)

Ocean: Gulf Stream Eddy Dynamics (MITgcm)
Gopalakrishnan et al., (2019)

River Transport: StreamFlow in WRF-Hydro
(Hurricane Florence)

Gharamti et al., (2021)
Earth System Observations (others available)

Instruments supported by RTTOV

- AIRS
- MODIS
- SSUSI
- MADIS
- CHAMP
- VTEC
- ATCF
- TPW
- DWL

Supporting Organizations:
- AURA
- GOES
- GPS
- AIRS
- MOPITT
- MPD
- Radar
- GMI
- GRACE
- COSMOS
- Atmosphere Fluxes
- NCEP + ACARS
- QuikSCAT
- CMEMS
- PODAAC
- NCDC
- ROMS
- WOD
- GTSPP
- CONAGUA
- SIF
- SIF
- SMOS, SMAP
- SNOTEL
- VTEC
- GSI2DART
- ok_mesonet
- ok_mesonet
• Observations combined with a model forecast to produce an improved forecast (‘analysis’).
• Typically adjust the system state, but also model parameters

Bayes Theorem

Posterior ~ Prior · Observation Likelihood

‘Update’ or ‘analysis’  Model generated  Earth System Observations

5 prior model estimates of temperature  1 new observation of temperature

This is an ‘observed’ state variable, but what about ‘unobserved’ state variables?
Basics of EnKF Data Assimilation

- Imagine you were modeling temperature across Salt Lake City but only had temperature observations at Alta Ski Resort

Ensemble of model generated temperatures

Apply correction to model with observed temp

Apply correction to unobserved temp

Generate posterior

- This is a simple example, but in complex ESMs this can be applied across entire model state: both in physical space, and across different variables.

- How can we apply correlations to improve model performance for Land DA?
Expanding Earth System Observations

Remote Sensing Satellites

2018

2019

ECOSTRESS
Studying Plant Water Use and Stress
On ISS

GEDI
ECOSYSTEM LIDAR

HISUI
HYPER SPECTRAL IMAGING
FROM SPACE

2020

Ground Based Ecological Observation Networks: NEON, Ameriflux

Stavros et al., (2017)

Metzger et al., (2019)

Land-Atmosphere CO₂, water exchange

Biological Measurements
Components of a land surface model (CLM)

**Carbon and nitrogen cycles**
- Photosynthesis
- BVOCs
- Autotrophic respiration
- Heterotrophic respiration
- Phenology
- Vegetation C/N
- Soil C/N
- N mineralization
- N uptake
- N fixation
- N leaching
- Denitrification
- N deposition
- CH$_4$
- N$_2$O
- Litterfall
- Root litter

**Hydrology**
- Precipitation
- Transpiration
- Evaporation
- Infiltration
- Sublimation
- Melt
- Throughfall
- Surface runoff
- Aquifer recharge
- Water table
- Saturated fraction
- Unconfined aquifer
- Sub-surface runoff

**Energy balance**
- Direct solar
- Diffuse solar
- Downwelling longwave
- Reflected solar
- Absorbed solar
- Emitted longwave
- Latent heat flux
- Sensible heat flux

Gross Primary Productivity (GPP)
Ecosystem Respiration (ER)
Net Ecosystem Production (NEP)
NEP = GPP – ER

- The carbon cycle is coupled to, and influenced by the nitrogen, water cycles and surface energy balance
Limitations of remotely-sensed land observations

Leaf Area, Biomass, SIF

Soil Moisture, Temp, Snow

- Spatial Coverage
- Temporal Coverage
- Sub-surface Coverage

Snow: up to 12 vertical layers
Ice, water etc.

Soil: up to 25 vertical layers
Carbon, water, ice properties

Snow (SWE)

Soil moisture, Carbon, Temp

Bedrock
Limitations of ground-based land observations

- Horizontal Spatial Correlations Important for limited surface observation network

Spatial Coverage
Temporal Coverage
Sub-surface Coverage

Snow: up to 12 vertical layers
Ice, water etc.

Soil: up to 25 vertical layers
Carbon, water, ice properties

Soil moisture, carbon, temp

Land
Atmosphere
Carbon/Water flux

NEON site

Soil moisture, Carbon, Temp

Bedrock
Carbon Monitoring Across Western US

US Drought Monitor, Oct 26, 2021

Palmer Drought Severity Index (1895-present for California)

Intensity
- None
- D0 (Abnormally Dry)
- D1 (Moderate Drought)
- D2 (Severe Drought)
- D3 (Extreme Drought)
- D4 (Exceptional Drought)

wetter
drier
• Complex terrain challenges traditional carbon monitoring, flux towers, atmospheric inversions

• Approaches to quantify regional land-atmosphere exchange of CO₂
Generating an assimilation in CLM5-DART

1) Initialize model states
2) Generate ensemble spread
3) Assimilate observations

- 'Inflation' helps maintain ensemble spread
  - Ensemble Sampling error
  - Model vs. Observation Bias

CAM DART Reanalysis (80 member ensemble)

CAM4 Reanalysis (~2°) → CAM6 Reanalysis (~1°)

Raeder et al., (2012, 2021)

Ds199.1 | DOI: 10.5065/38ED-RZ08
Ds345.0 | DOI: 10.5065/JG1E-8525
CLM5-DART Overview

CAM4 DART Reanalysis (80 member ensemble)

Grid Cell (~1°x1°)

‘Localized’ the adjustments to biomass: 7 carbon and 7 nitrogen state variables
Observations reduce biomass/leaf area, net carbon flux steady

- ~30% reduction in AGB and LAI respectively

<table>
<thead>
<tr>
<th>Simulation Name</th>
<th>AGB (kgC m(^{-2}))</th>
<th>LAI (m m(^{-2}))</th>
<th>GPP (gC m(^{-2}) month(^{-1}))</th>
<th>ER (gC m(^{-2}) month(^{-1}))</th>
<th>NEP (gC m(^{-2}) month(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free</td>
<td>1.98</td>
<td>1.31</td>
<td>48.18</td>
<td>47.18</td>
<td>1.00</td>
</tr>
<tr>
<td>CLM5-DART</td>
<td>1.36</td>
<td>0.96</td>
<td>38.49</td>
<td>37.21</td>
<td>1.28</td>
</tr>
</tbody>
</table>
Diagnostics of LAI/AGB observation acceptance and RMSE

**Leaf Area**: steady acceptance rate (90%) seasonal dependence, RMSE steady

**Biomass**: increasing acceptance rate (75%), decreasing RMSE

- Observations possible
- Observations assimilated
- Prior RMSE
Behavior for dominant Plant Functional Types

![Image of graphs showing Leaf Area (m² m⁻²) and Biomass (gC m⁻²) for Temp. Evergreen Forest, Boreal Evergreen Forest, C3 Grass, and Temp. Shrub over the years 1998 to 2010.]
CLM5-DART simulates weak carbon sink compared to FLUXCOM

- CLM5-DART (red) reduces biomass states create offsetting reductions in GPP and ER compared to free run.
CLM5-DART simulates weak carbon sink compared to FLUXCOM

- **CLM5-DART** (red) reduces biomass states create offsetting reductions in GPP and ER compared to free run

- **FLUXCOM** (yellow): Machine learning approach that trains satellite data and meteorology to flux tower data to generate a carbon flux product Jung et al., (2020).

**CLM5-DART:**
- **Strength:** more explicit disturbance history, not dependent on flux tower CO₂ data
- **Weakness:** limited adjusted variables (biomass)
CLM5-DART simulates weak carbon sink compared to FLUXCOM

1998-2011 Average Fluxes

Free CLM5

CLM5-DART

FLUXCOM

Weak, neutral uptake

Strong uptake
Water limitation shapes carbon uptake pattern

Soil moisture limitation (0-1)

GPP (gC m$^{-2}$ mth$^{-1}$)

Snow water equivalent (mm)

- Soil moisture limitation and GPP highly correlated (spring: $R=0.64$; summer: $R=0.67$)

- Simulated snow has low bias
Current Land Data Assimilation: Arctic

Arctic Boreal Domain (ABoVE Project), Led by: Xueli Huo, Andy Fox

Leaf Area Index (LAI)

Observation 2012-2019 average

Free 2012-2019 average

Assim 2012-2019 average

• 30% reduction in Leaf Area
Arctic Boreal Domain (ABoVE Project), Led by: Xueli Huo, Andy Fox

Aboveground Biomass (AGB) (gC m$^{-2}$)

- 70 % reduction in AGB
Current Land Data Assimilation: Soil Moisture

- Gap-Filling Soil moisture products across China
- European Space Agency Climate Change Initiative Essential Climate Variable (ECV)

ECV Soil Moisture Product (m$^3$ m$^{-3}$)

CLM-DART

Compares favorably to GLEAM Soil Moisture Data Product (1998)

Unbiased RMSD (m$^3$ m$^{-3}$)

Correlation

Led by: Daniel Hagan, Nanjing University of Information Science & Technology
Challenges in Land DA: Soil Moisture

- Soil moisture data are prone to systemic bias in magnitude

- Model/Data product bias is challenging to address

- The trends and patterns in the data are useful. Cumulative Distribution Function (CDF) matching re-scales data products to match the magnitude and variation of model

volumetric soil moisture (mm³/mm³); Koster et al., 2009

(Model) – (Data Product), Before

CDF Matching

(Model) – (Data Product), After

Reichle & Koster 2004 (GRL)
Current challenges in Land DA: Snow

Snow Hydrology: Snow Water Equivalent
Ice content
Water content

Snow Albedo: Surface Energy Balance
Black/organic carbon
Dust
Snow Grain radius

• CLM snow will compact and subdivide into layers depending upon layer thickness
• This creates unique snow properties for each layer
• This presents challenges for DA systems
Current challenges in Land DA: Snow

- Standard implementation of DART regression and update step will not work if layer (and property) does not exist for all ensemble members.
Current challenges in Land DA: Snow

- Standard Approach
- Snow (SWE) Observations
- Added Snow repartitioning algorithm

Model Estimated SWE

Snow Layer Property \( i = n \)

\[ \Delta \text{Total SWE} \neq \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Ice} \neq \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Liquid} \neq \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Depth} \neq \Sigma (\Delta \text{Layers}) \]

Snow updates are not internally consistent

Repartitioning Algorithm

\[ \Delta \text{Total SWE} = \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Ice} = \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Liquid} = \Sigma (\Delta \text{Layers}) \]
\[ \Delta \text{Total Depth} = \Sigma (\Delta \text{Layers}) \]

Snow updates are internally consistent
Challenges in Land DA: Solar-Induced Fluorescence

- SIF is a useful indicator of timing/magnitude of photosynthesis (GPP)

- Strong SIF-GPP relationship across many vegetation types

Sun et al., (2018)

Magney et al., (2019)
Advancing observations & models together

Smith et al., (2020)

Expanding satellite observation network

CLM 4.5
(Soil Moisture Stress Formulation)

Current: CLM 5.0
Added Hydraulic Stress & SIF

Increasing model complexity

Leaf water potential $\Psi_{\text{leaf}}$
SIF$_{\text{canopy}}$

SIF$_{\text{leaf}}$
Advances in DART

- Increased emphasis on coupled Earth System assimilations (e.g. land-atmosphere coupling)

- Addressing Bounded Quantities:
  General Ensemble Filtering Framework Using Quantiles (GEFFQ) – Jeff Anderson
For more information:

https://dart.ucar.edu
https://docs.dart.ucar.edu
dart@ucar.edu

This research was supported by the NASA CMS Program (awards NNX16AP33G and 80NSSC20K0010). CESM is sponsored by the National Science Foundation and the U.S. Department of Energy. We would like to thank the Center for High Performance Computing at the University of Utah. We would also like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation, through allocation awards UUSL0005 and UUSL0007.
Example of DART workflow

CAM4 DART Reanalysis
(80 member ensemble)

Ds199.1 | DOI: 10.5065/38ED-RZ08

Observations

Aboveground Biomass (gC m⁻²)
Liu et al., (2015)

Leaf Area (m² m⁻²)
Zhu et al., (2015)

CLM 5 terrestrial biosphere model

Initial Prior State

Fluxes: NEP + GPP + Δ ER + Δ ET + Δ

Biomass Leaf Area + Δ

New Posterior State

Fluxes: NEP + Δ GPP + Δ ER + Δ ET + Δ

Biomass + Δ

Leaf Area + Δ

Soil carbon + Δ

Soil water + Δ

Research Article | Open Access

Improving CLM5.0 Biomass and Carbon Exchange Across the Western United States Using a Data Assimilation System

Brett Racza, Timothy J. Hoar, Henrique F. Duarte, Andrew M. Fox, Jeffrey L. Anderson, David R. Bowling, John C. Lin,

First published: 19 June 2021 | https://doi.org/10.1029/2020MS002421
Assimilating Surface Soil Moisture Observations (Passive/Active Microwave Bands)
Led by: Daniel Hagan, Nanjing University of Information Science & Technology

LPRM_SOIL_MOISTURE
"observation"
Current Land Data Assimilation: Arctic

Arctic Boreal Domain (ABoVE Project)
Led by: Xueli Huo, Andy Fox and others

Leaf Area (Monthly)

The mean annual LAI in the assimilation run decreased by 63.7% compared with the free run.
Future Directions

Additional data streams help constrain carbon cycling

Using high res land cover maps for improved forward operators (PFT specific).

Finer Spatial Resolution?

Atmosphere:

- CAM4 Reanalysis (~2°)
  - DOI: 10.5065/38ED-RZ08

- CAM6 Reanalysis (~1°)
  - DOI: 10.5065/JG1E-8525

Parameter Estimation

Land surface:

CLM parameters
“Meeting in the middle manuscript”

Alexei Shiklomanov

→ Soil moisture/vegetation optical depth/radiative transfer characteristics
For leaf properties←

Add SIF here as well leaf to canopy level SIF getting closer to observations