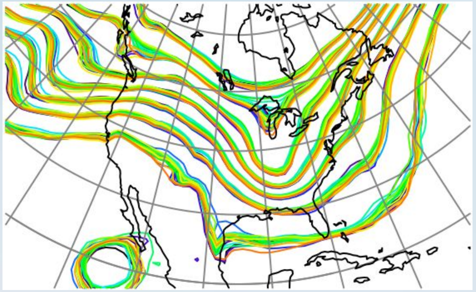


CESM Parameter Estimation with the Data Assimilation Research Testbed (DART)

Jeff Anderson representing
CISL/Data Assimilation Research Section

CESM Parameter Estimation: 12 June 2023

NCAR | National Center for
UCAR | Atmospheric Research

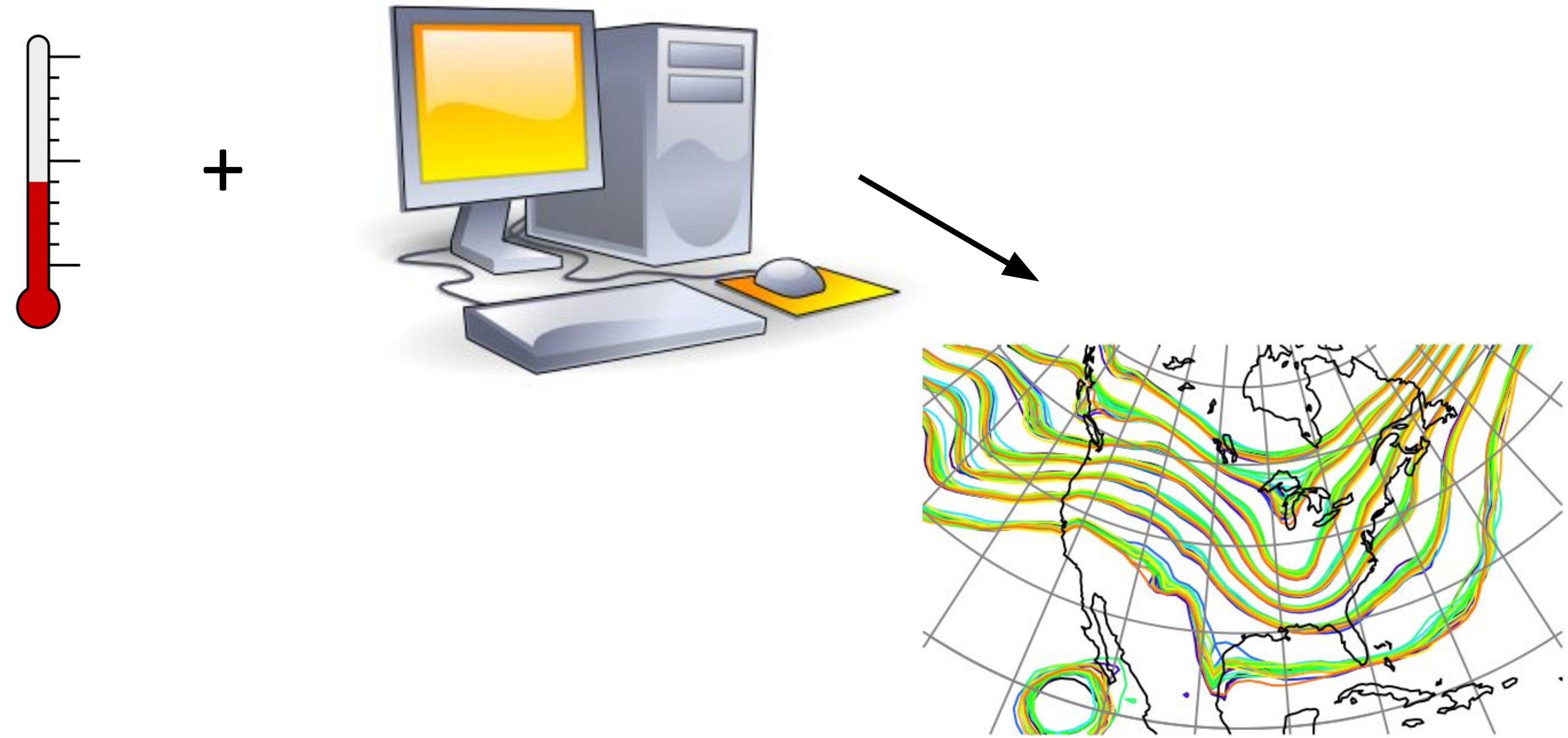
The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NSF

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

What is ensemble data assimilation?

Observations combined with an ensemble of model forecasts…



Data Assimilation Research Testbed (DART)

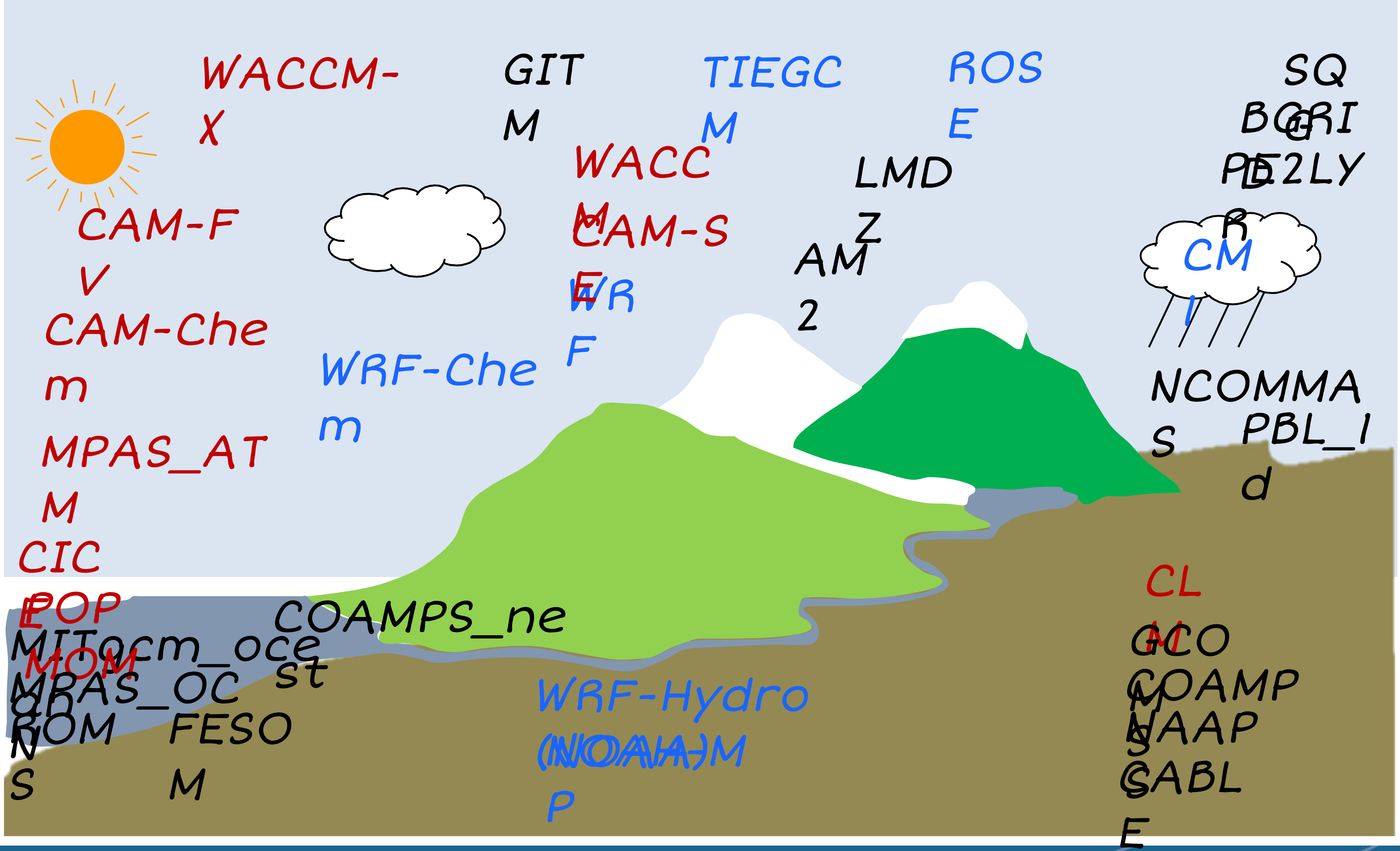
- A state-of-the-art ensemble Data Assimilation System for Geoscience.

- DART can be used for observing system simulation experiments:
  - Evaluate the impact of existing or proposed (remote sensing) observations,
  - Design observing systems.

- DART can also be used for real assimilation.
  - New remote sensing observations can be assimilated with any appropriate model.

- State-of-the-art ensemble DA is essential to study impact of novel observations on prediction, predictability, and model development.

Geophysical Models Interfaced to DART

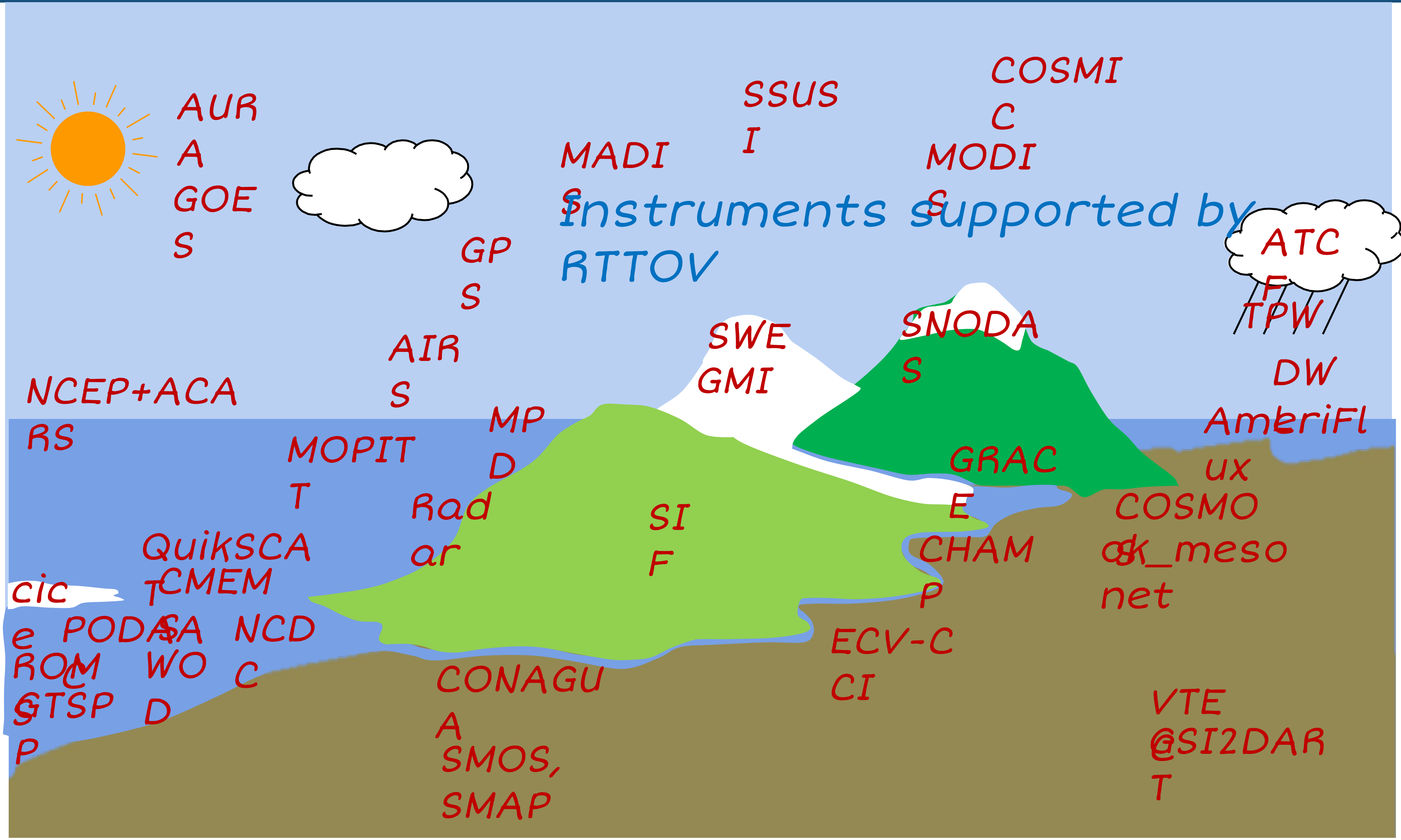


CESM Models Red

NCAR Models Blue

Selected other models Black

Earth System Observations (others available)



DART CAM Reanalysis demonstrates maturity

The CAM6+DART Reanalysis for Earth System Science

The Earth system can be viewed as distinct but connected components: atmosphere, land, ocean, cryosphere, biosphere, et cetera (Fig. 1). Data assimilation can help us create the best available representation of the state of Earth, but it requires relevant observations and a forecast model which represents all of the components of interest.

Earth system components interact in many ways at the interfaces between them.

NCAR's Community Earth System Model (CESM) can run forecasts with a flexible choice of "active" components, in which the component model state evolves according to equations, and "data" components, in which the component state is read from a data file. For example, to generate atmospheric forecasts, the configuration could have active atmosphere and land components, but simply read sea surface temperatures (SSTs) from data files, instead of running an expensive, active, ocean component to generate SSTs. CESM has been developed at NCAR for decades, and has evolved to work effectively with DART through the efforts of the CESM Software Engineering Group (M. Vertenstein, S. Goldhaber, J. Edwards) and R. Montuoro.

Data assimilation has been extensively applied to the atmosphere for decades, but not to the surface components until more recently. One hurdle has been that the surface components tend to be more slowly varying, so they require atmospheric forcing over long time spans. It's expensive to run an atmospheric model, and many experiments may require the same atmospheric forcing, which would be wasteful to regenerate each time. Further, research shows that an ensemble of surface models requires an ensemble of forcing from the atmosphere in order to maintain the necessary ensemble spread (Fig. 2).

So the cost is multiplied by the size of the ensemble. There are thus compelling reasons to generate an ensemble of atmospheric forcing once and archive it for repeated use.

To satisfy this need the DART team has generated a "reanalysis" spanning years 2011-2020 using DART, CESM (v2.1) with an active atmospheric model component, the Community Atmosphere Model version 6(CAM6), and several million observations per day. This reanalysis shares characteristics with widely used reanalyses, such as ERA5 (Hersbach, et al., 2020), JRA55 (Kobayashi, et al., 2015), and MERRA-2 (Gelaro, et al., 2017). The primary goals of those are to provide a high resolution (spatial and temporal) description of the

Figure 1: Earth system components interact in many ways at the interfaces between them.

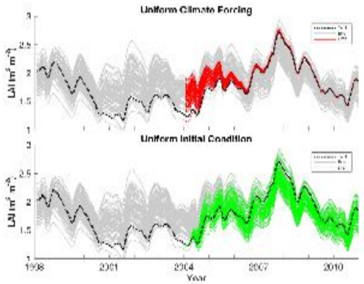


Figure 2: A single atmospheric forcing allows an ensemble to collapse (top). Multiple atmospheric forcings cause the spread to increase (bottom). Picture courtesy of A. Fox.

10 Year Reanalysis,

80 Members,

Provides forcing ensembles for other CESM models,

Available from NCAR RDA.

https://rda.ucar.edu/datasets/ds345.0/

DART CAM Reanalysis demonstrates maturity

The CAM6+DART Reanalysis for Earth System Science

The Earth system can be viewed as distinct but connected components: atmosphere, land, ocean, cryosphere, biosphere, et cetera (Fig. 1). Data assimilation can help us create the best available representation of the state of Earth, but it requires relevant observations and a forecast model which represents all of the components of interest.

Earth system components interact in many ways at the interfaces between them.

NCAR's Community Earth System Model (CESM) can run forecasts with a flexible choice of "active" components, in which the component model state evolves according to equations, and "data" components, in which the component state is read from a data file. For example, to generate atmospheric forecasts, the configuration could have active atmosphere and land components, but simply read sea surface temperatures (SSTs) from data files, instead of running an expensive, active, ocean component to generate SSTs. CESM has been developed at NCAR for decades, and has evolved to work effectively with DART through the efforts of the CESM Software Engineering Group (M. Vertenstein, S. Goldhaber, J. Edwards) and R. Montuoro.

Data assimilation has been extensively applied to the atmosphere fo[...] the surface components until more recently. One hurdle has [...] components tend to be more slowly varying, so the[...] time spans. It's expensive to run an atmosp[...] require the same atmospheric forcing, wh[...] time. Further, research shows that an e[...]semble of forcing from the atmosphere i[...]ble spread (Fig. 2).

S[...]semble. There are thus compelling reasons to g[...]ric forcing once and archive it for repeated use.

To [...]e DART team has generated a "reanalysis" spanning years 2011 2020 usin[...] CESM (v2.1) with an active atmospheric model component, the Community Atmosphere Model version 6(CAM6), and several million observations per day. This reanalysis shares characteristics with widely used reanalyses, such as ERA5 (Hersbach, et al., 2020), JRA55 (Kobayashi, et al., 2015), and MERRA-2 (Gelaro, et al., 2017). The primary goals of those are to provide a high resolution (spatial and temporal) description of the

Figure 1: Earth system comp[...] many ways at the i[...]

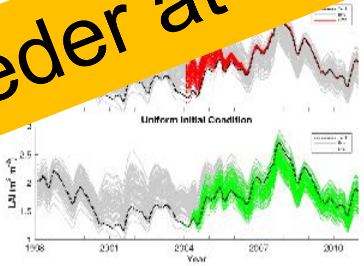


Figure 2: A single atmospheric forcing allows an ensemble to collapse (top). Multiple atmospheric forcings cause the spread to increase (bottom). Picture courtesy of A. Fox.

Talk to Kevin Raeder at his poster for more information.

10 Year Reanalysis

80 Mem

vides forcing ensembles for other CESM models,

Available from NCAR RDA.

https://rda.ucar.edu/datasets/ds345.0/

Parameter estimation with state augmentation

Normally, assimilation uses observations to update model state variables

Things like gridded T, U, V, Q, leaf carbon, phytoplankton concentration, ice area…

If each ensemble forecast is generated with its own value of a parameter:

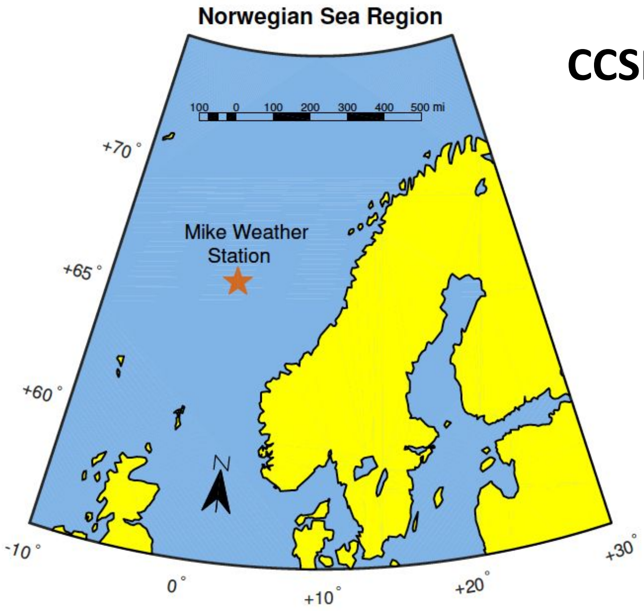
- The estimates of the parameter can be updated by observations
- Called state augmentation

Challenges:

- Parameters (by definition) do not change during forecasts
- Have to generate a prior ('guess') ensemble of values
- Never directly observed (state variables aren't either)
- Constraints on legal (reasonable) values

Despite challenges, lots of success!

Ocean biogeochemical parameter estimation



Study location (station M) in the Norwegian Sea

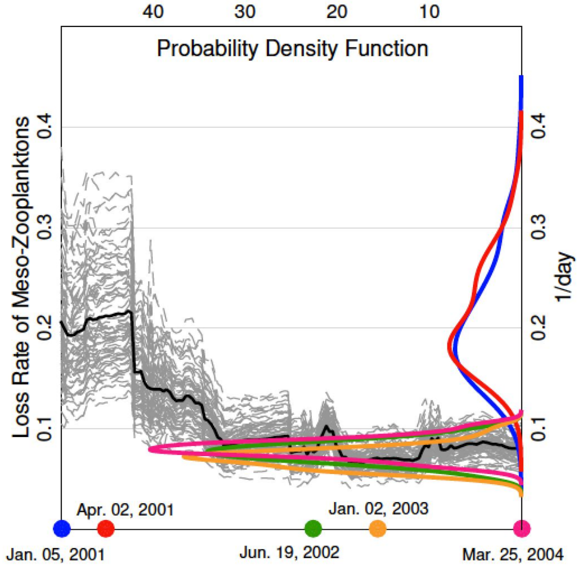
CCSM4: Community Climate System Model (Gent et al., 2011) with ocean components:

- ❑ Physics: Miami Isopycnic Coordinate Ocean Model (MICOM, Bleck et al., 1992)
- ❑ Biogeochemistry: The HAMburg Ocean Carbon Cycle (HAMOOC5, Maier-Reimer et al. 2005)
    - ▪ NPZD-type BGC model with Carbon

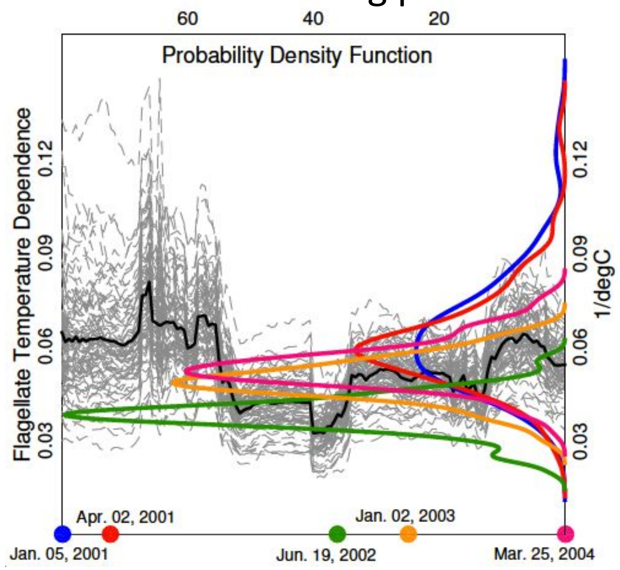
1D State-Parameters Estimation using the EnKF:

- ❑ Weekly updates
- ❑ BGC state: 15 variables (tracers) & 11 parameters
- ❑ Data includes nutrient& oxygen profiles, surface chlorophyll and partial pressure of CO_2

Example of a parameter that converged



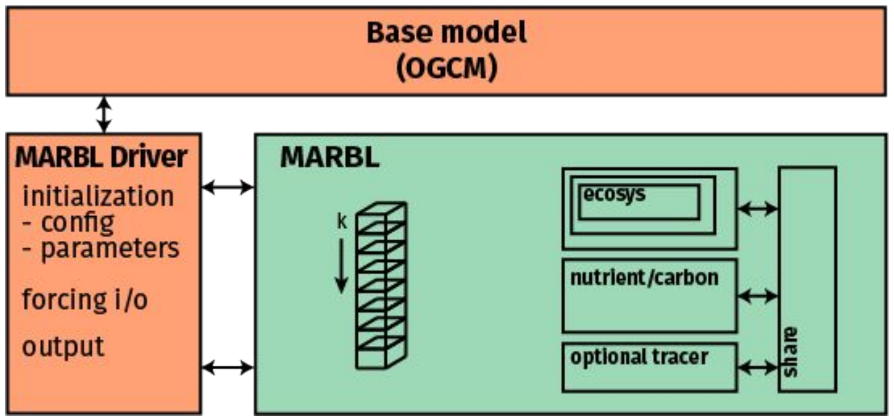
Example of a parameter that that shows a seasonal cycle closely related to the blooming periods



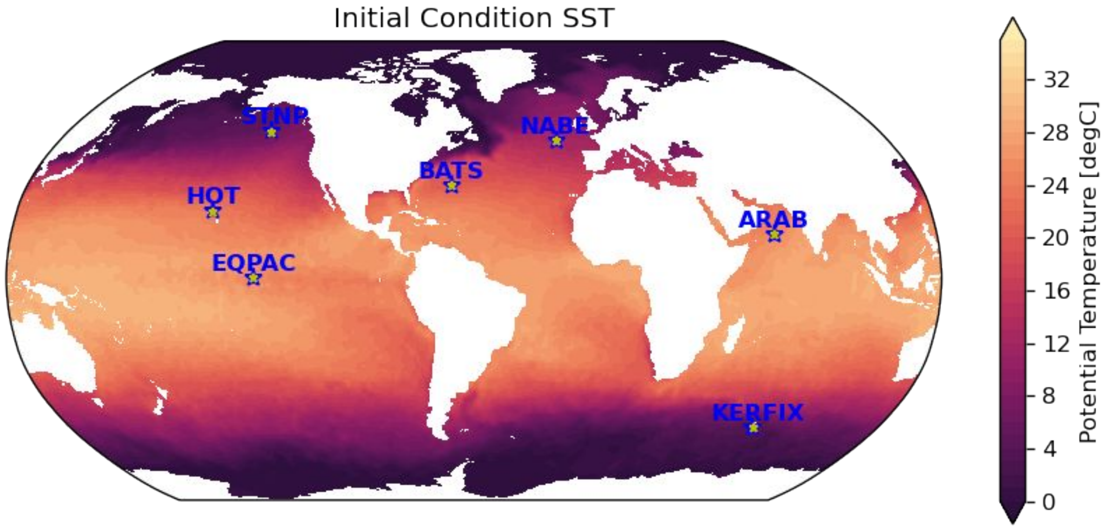
Gharamti, M., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., & Bertino, L. (2107). Ensemble Data Assimilation for Ocean Biogeochemical State and Parameter Estimation at Different Sites. Ocean Modelling, 112, 65-89.

Ocean biogeochemical parameter estimation in MARBL

- Objective: Implement 1D model configuration to optimize a set of ocean parameters at testbed sites using DART

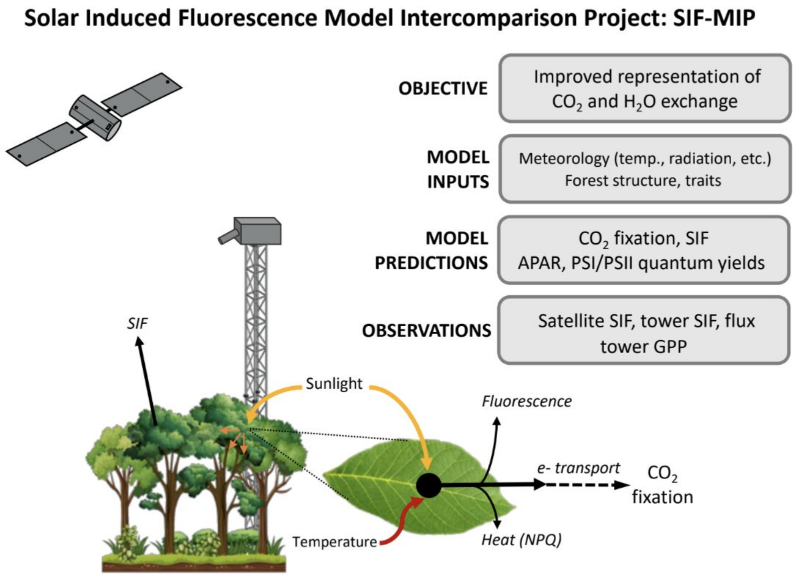


- Test Locations Identified



Robin Armstrong (SIParCS), Dan Amrhein, Moha Gharamti and others

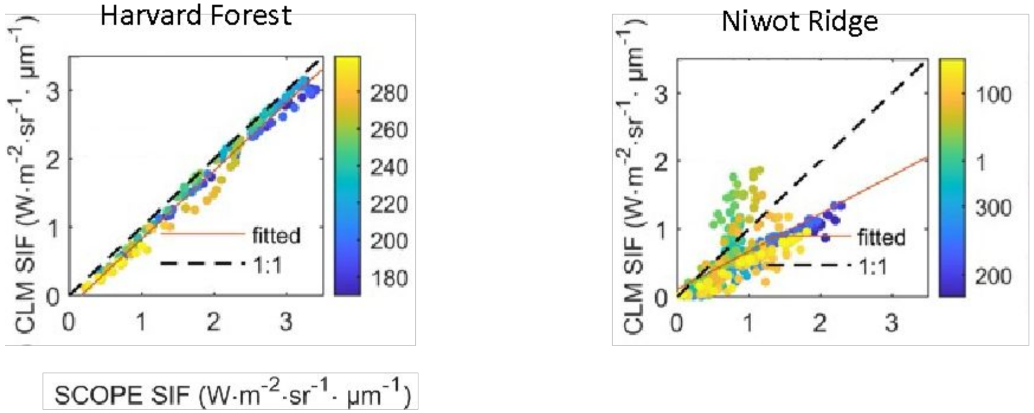
Land parameter estimation in CLM: SIF-MIP2



Provided by N. Parazoo

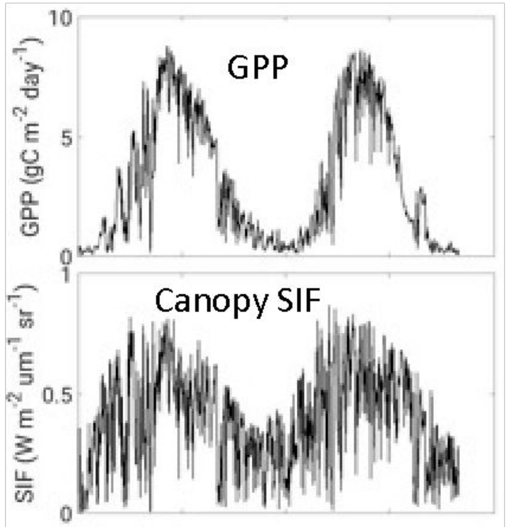
- Group 1: Free simulations (~10 models)
- Group 2: Data Assimilation (ORCHIDAS, CLM-DART, CARDAMOM)
- Group 3: Radiative Transfer Models

1) Create SIF forward operator (leaf to canopy)



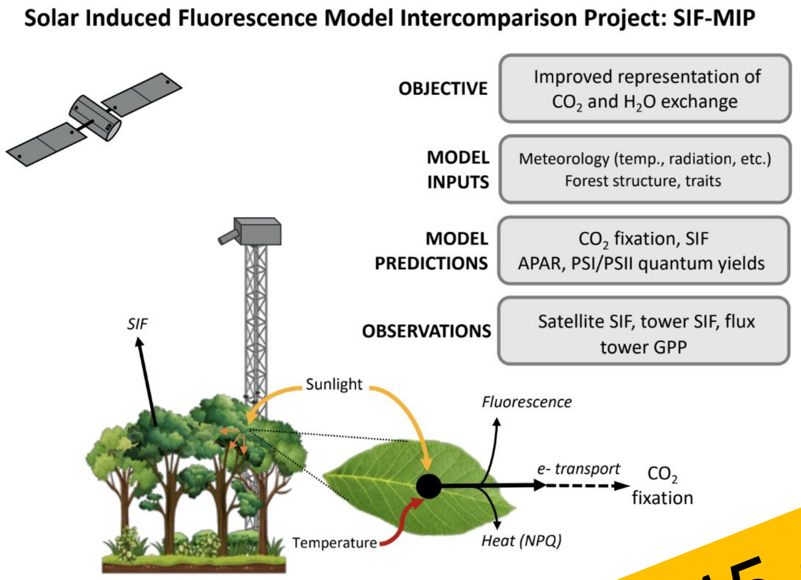
R. Li et al., (2022)

2) CLM performance (SIF-MIP2)



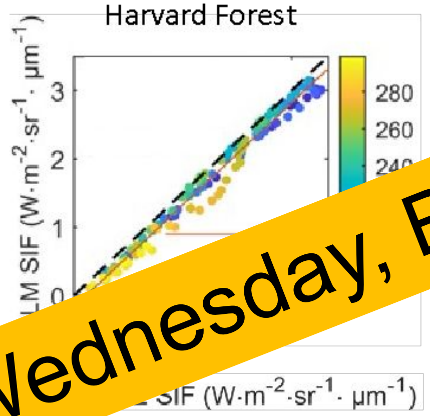
Niwot Ridge,
Years: 2017-18

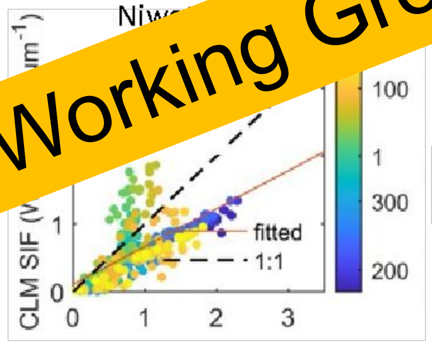
Land parameter estimation in CLM: SIF-MIP2



Provid

1) Create SIF forward operator (leaf to canopy)



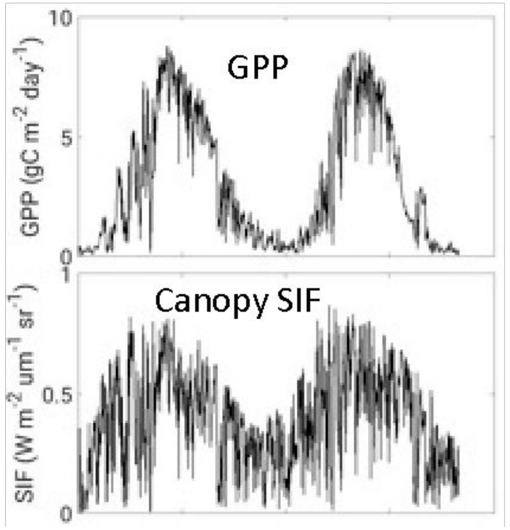


R. Li et al., (2022)

Brett Raczka, 10:45 am Wednesday, BGC Working Group

- Group 1 … ations (~10 models)
- … Assimilation
  … IDAS, CLM-DART, CARDAMOM)
- Group 3: Radiative Transfer Models

2) CLM performance (SIF-MIP2)



Niwot Ridge,
Years: 2017-18

Estimating Carbon Sources Using CAM

Development and Applications of a Carbon-Weather Data Assimilation System

by

Stephanie M. Wuerth

Doctor of Philosophy in Earth and Planetary Science
Designated Emphasis in Computational and Data Science and Engineering

University of California, Berkeley

Professor Inez Y. Fung, Chair

This dissertation explores the utility of high-resolution satellite carbon dioxide (CO_2) and water vapor measurements for advancing climate treaty verification, for improving numerical weather prediction (NWP), and for understanding natural carbon cycling in the terrestrial biosphere. We present a series of Observing System Simulation Experiments (OSSEs) using a carbon-weather data assimilation (DA) system, where the state vector comprises weather

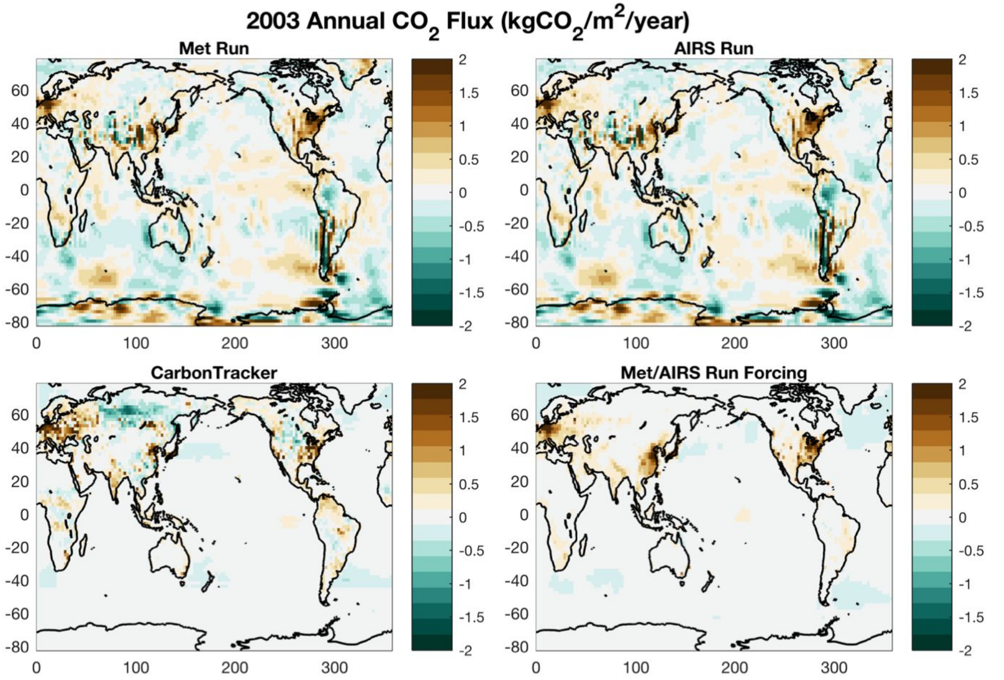


Figure 4.1: Annual CO_2 surface fluxes for the year 2003, in kg CO_2/m^2/year. The upper left panel shows the flux calculated using equation 4.2 from Met-run fields ($\overline{\Phi_{Met}}$). The upper right panel shows the same using AIRS-run fields ($\overline{\Phi_{AIRS}}$). The lower left panel is the annual total flux from CarbonTracker (Φ_{CT}), and the lower right panel is the annual flux used as forcing in both the Met- and AIRS-runs (Φ_{Prior}).

Gravity Wave Drag Efficiency Parameter Estimation in CAM

Data Assimilation and Parameter Estimation in CAM3.0
Using the Data Assimilation Research Testbed
Kevin Raeder, Jeff Anderson, Tim Hoar, Hui Liu
Report to AMWG Working Group
2 March, 2005

1. Overview of DART.
2. Assimilation / Prediction (NWP) with CAM 3.0.
3. Parameter estimation using data assimilation.
4. Synthetic observation assimilation for predictability / diagnosis.

Gravity Wave Drag Efficiency Parameter Estimation in CAM

Efficiency factor for orographic gravity wave drag generation is treated as an additional 2D state variable and assimilated.

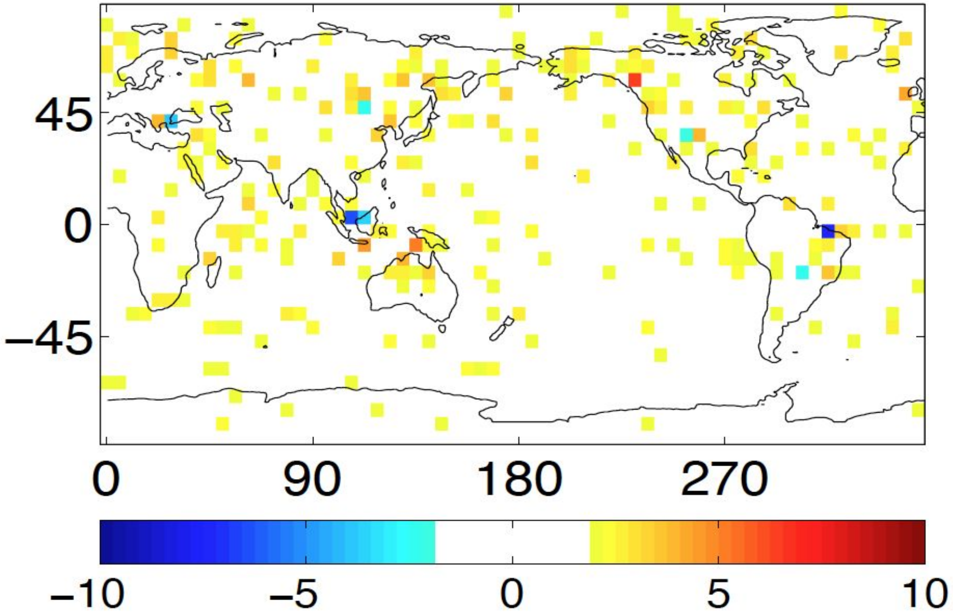
Model: CAM 3.0.7 T21L26.

Assimilation / Prediction Experiments:

- 80 member ensemble divided into 4 equal groups.
- Initialized from a climatological distribution (huge spread).
- Initial tests for January, 2003.
- Uses most observations used in reanalysis.
- Assimilated every 6 hours; +/- 1.5 hour window for obs.

Gravity Wave Drag Efficiency Parameter Estimation in CAM

Climate Model Parameter Estimation via Ensemble Data Assimilation



T21 CAM assimilation of gravity wave drag efficiency parameter.

Oceanic values are noise (should be 0).

0< efficiency< ~4 suggested by modelers.

Positive values over NH land expected.
Problem: large negative values over tropical land near convection.
May reduce wind bias in tropical troposphere, but for 'Wrong Reason'.

Fixing lack of cumulus momentum drag?

<u>Assimilation tries to use free parameter to fix ALL model problems</u>

The long lost g whiz slide

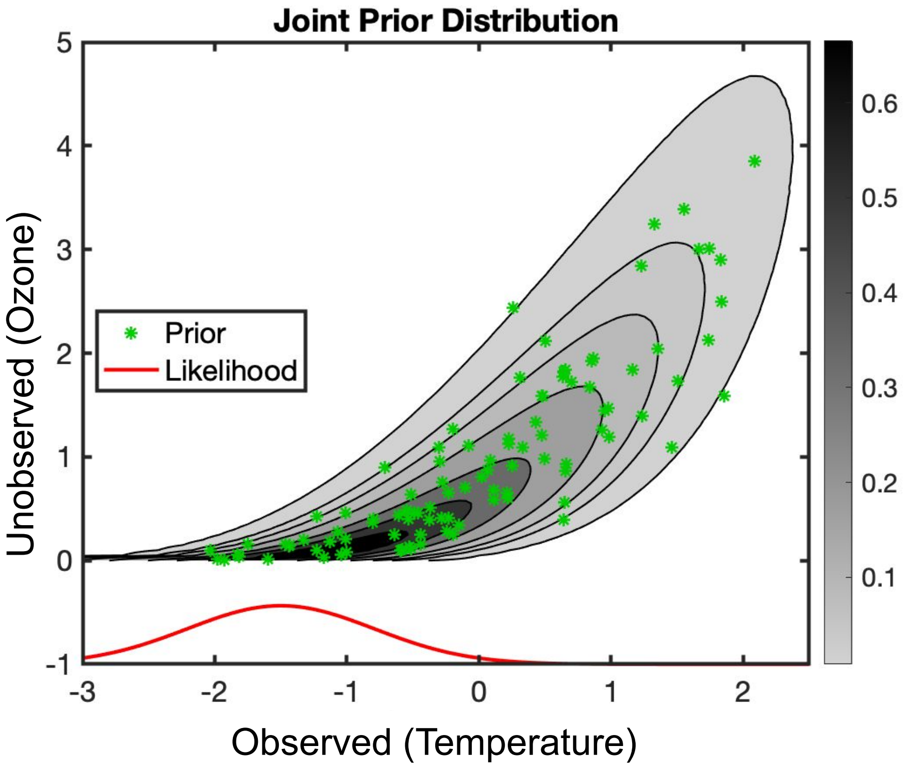
What planet are we really on?

Estimating earth's gravity with DART and CAM.

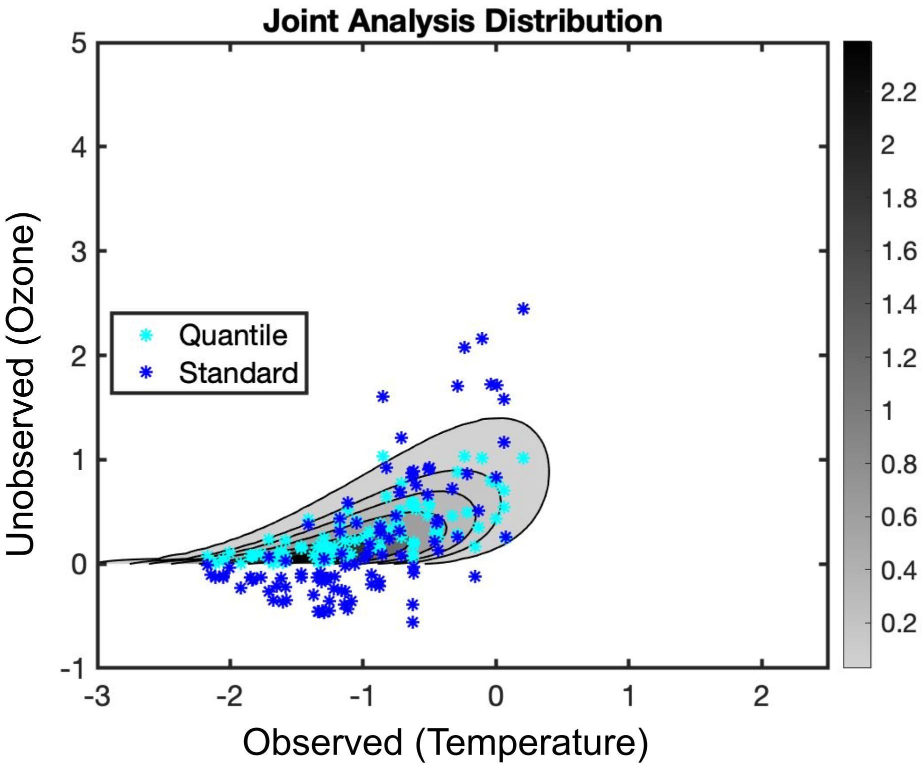
Don't expect to get the right answer.

DART: Novel, General Solutions for Nonlinear, Non-Gaussian Problems

Prior for normal-gamma distribution with 100 member ensemble.

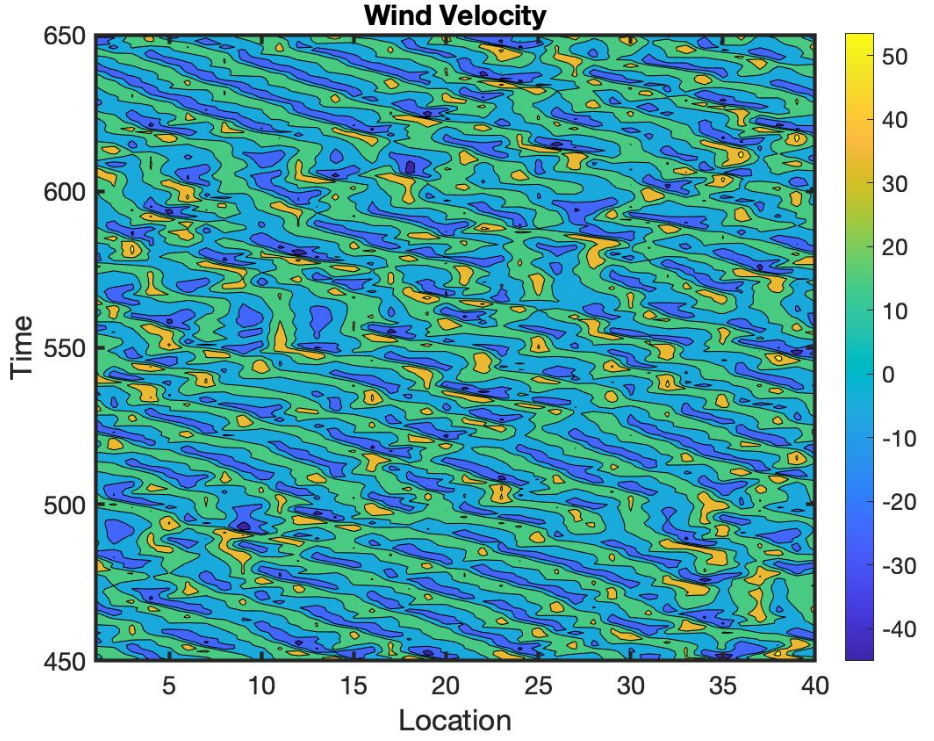


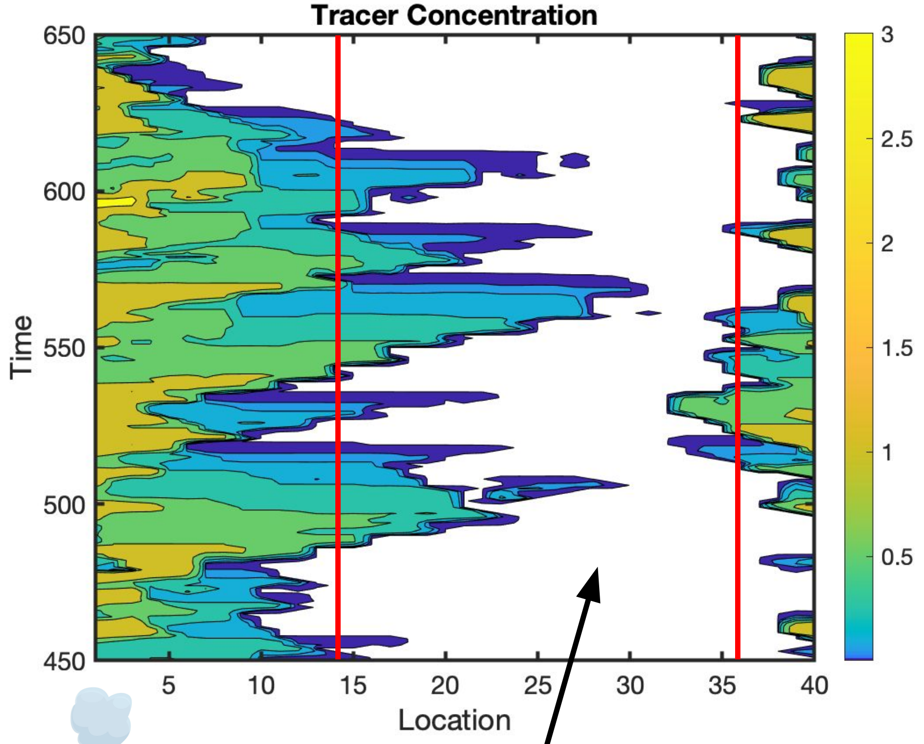
Bounds enforced. Nonlinear aspect respected.



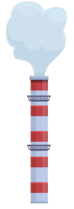
Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





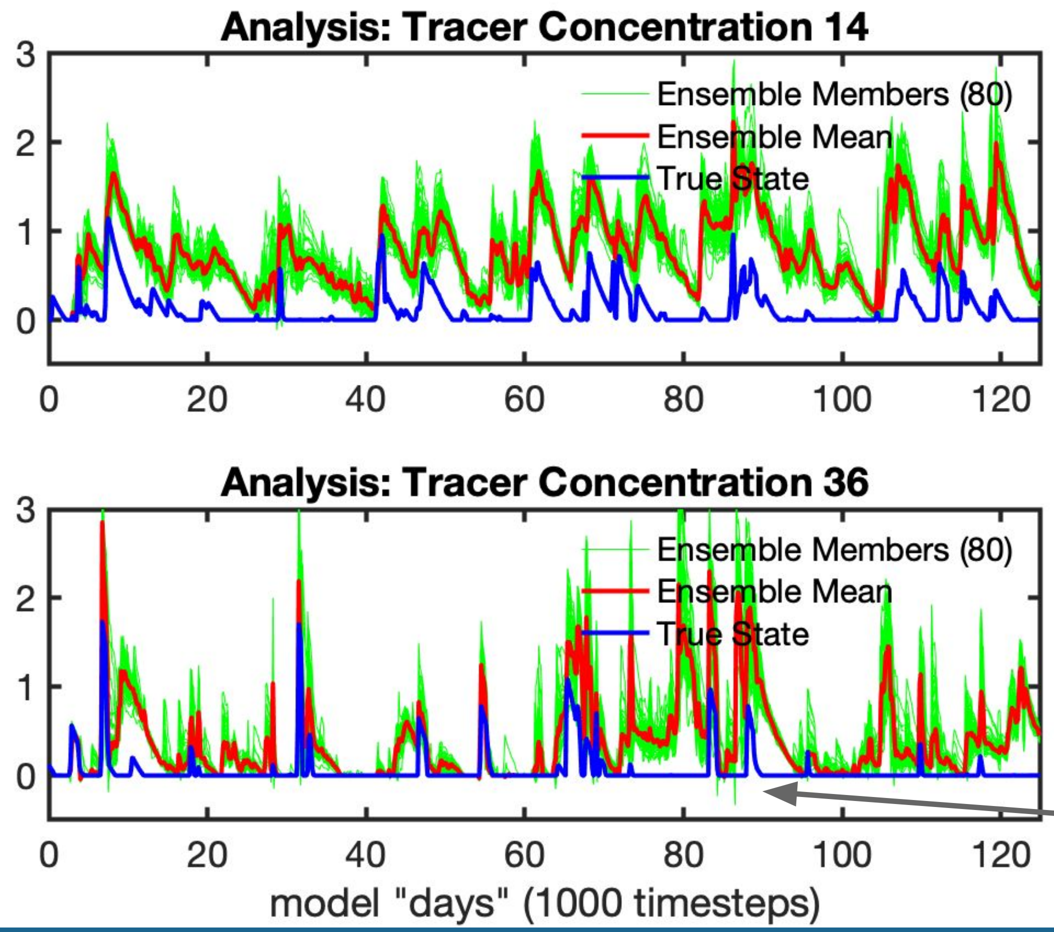
Constant Source.



Concentration can be zero far from source.

Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.



Observe state and concentration infrequently at each point.

Concentration error is truncated normal.

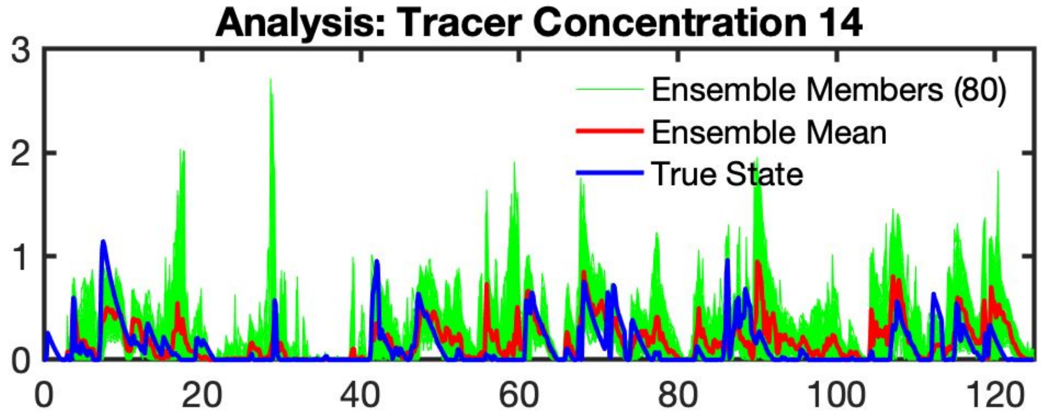
EAKF (linear, Gaussian) has large bias for tracers.

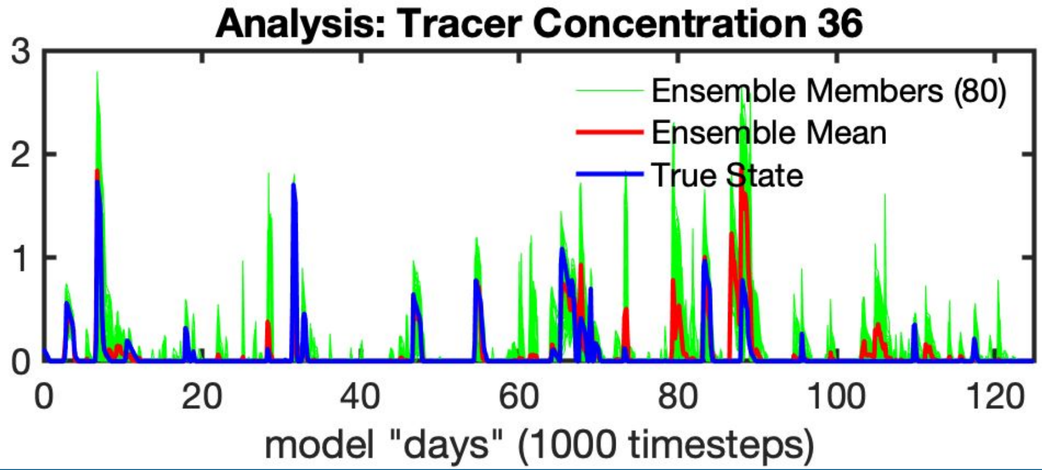
Can't go to all zeros.

Some negative values.

Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





Observe state and concentration infrequently at each point.

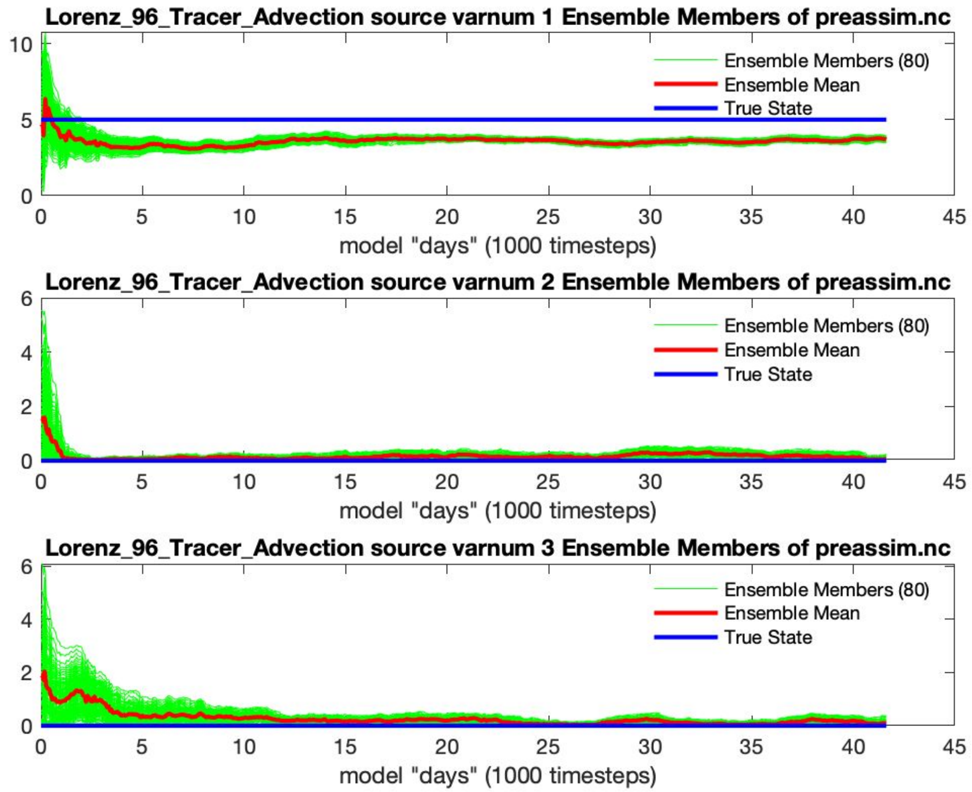
Concentration error is truncated normal.

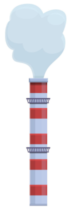
New method is (nearly) unbiased.

Can go to all zeros.

No negative values.

Low-Order Tracer Advection Model: Source Estimation





If sources are unknown, can also estimate them.
Example: single time constant source at point 1.

Systematic error becomes a huge problem.

Prior choices bias results.

These problems are revealed by DA,
but apply to any parameter estimation
method.

Summary: Parameter Estimation in CESM

DART can provide parameter estimation for most CESM models (with some additional coding)

New DA advances can improve estimates (including uncertainty)

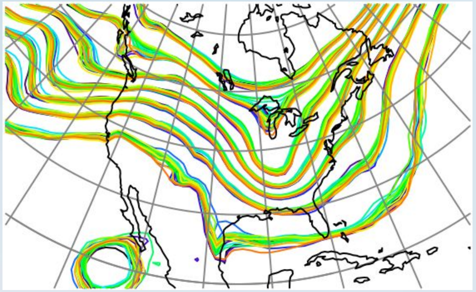
Parameter estimation can be effective when used carefully with:

- Methods for avoiding ensemble collapse (if desired)
- Appropriately constructed prior ensembles
- Appropriate parameter bounds

All parameter estimation methods:

- Can only answer specific questions
- Cannot estimate ‘true’ values of single parameters when others are wrong
- Cannot escape the information limits of model and observation errors

DART Webpage: dart.ucar.edu
Email the DART team: dart@ucar.edu



Questions?

The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

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