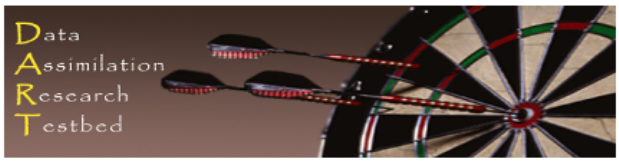
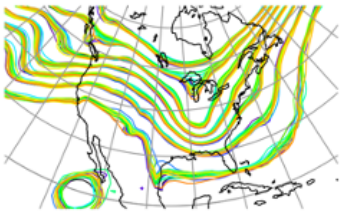


A New CAM6+DART 80-Member Ensemble Reanalysis for 2011-2019: Initial Conditions for CAM, CLM and CICE Forecasts and Forcing for POP, CLM and CICE



©UCAR 2019

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.

NCAR | National Center for
UCAR | Atmospheric Research

Thanks for the opportunity to tell you about this new data set.
It's an honor ot be the first presentation of the first WG meeting.
We believe that it will be very useful in several ways
to people who are interested in using CESM
for any hindcasts shorter than climate time scales.

0:20

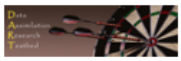
Context and Goals

This new reanalysis is similar in quality to the NCEP/NCAR reanalysis
but with 80 ensemble members in addition to the mean.
It's also similar to the 20th Century Reanalysis.

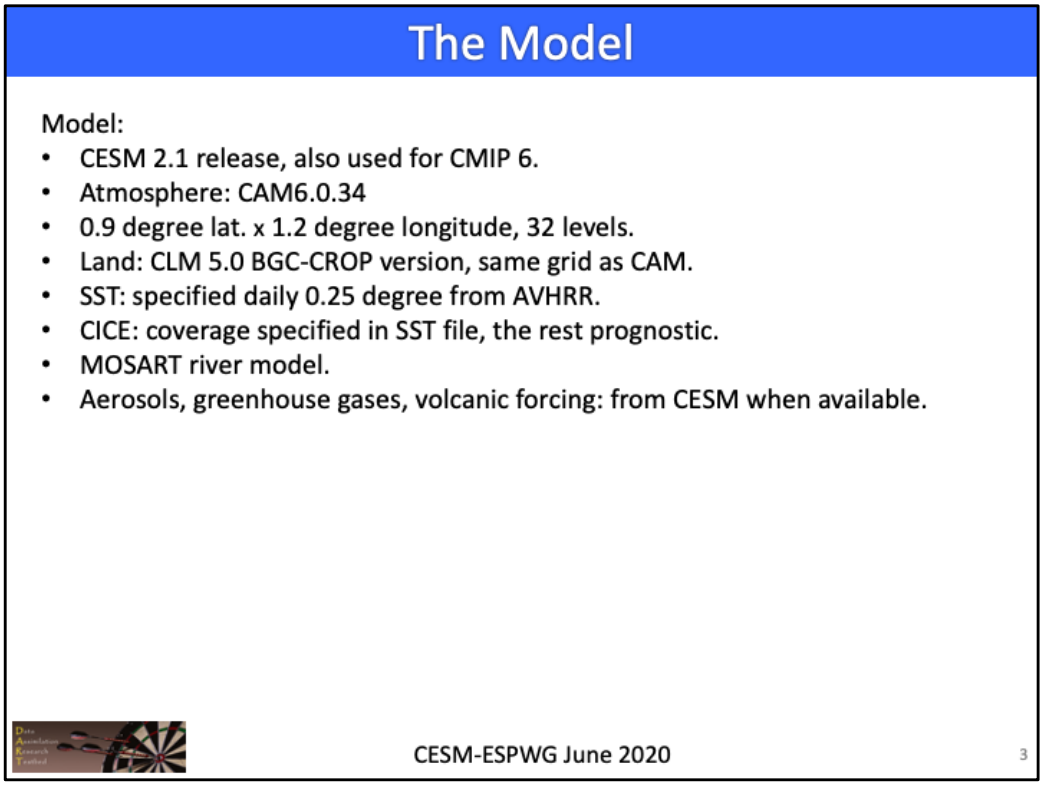
It's an upgraded version of the CAM4, 2-degree reanalysis finished in 2012:
+ CAM6 @ 1-degree,
+ more data products for use in hindcasts.

Most of the data products are native to CESM:
+ No initial shock from importing ICs from a foreign model.
+ Data atmosphere files fed directly to the coupler for any DATM application.
+ Initial conditions spun up to the actual weather for CAM, plus CLM, CICE, and MOSART

Estimates of uncertainty are built-in:
~ Ensemble spread shows model uncertainty
~ DART diagnostics show assimilation quality



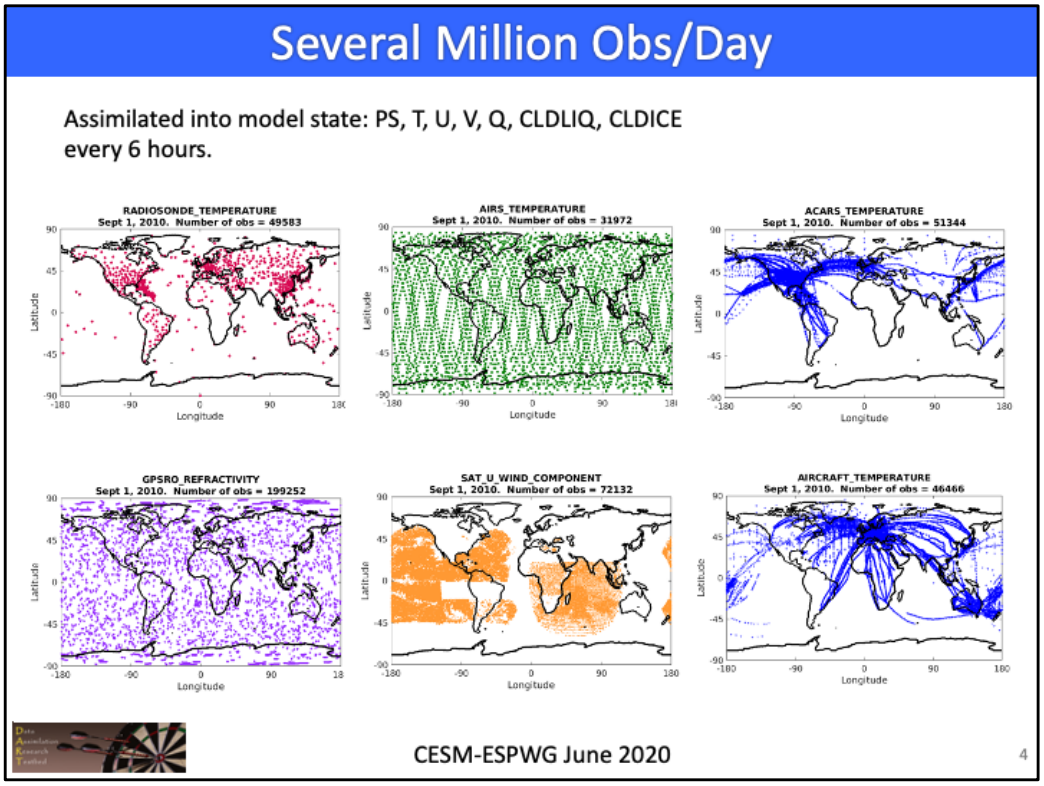
1:00



Assimilations have 3 components: a model, observations, and DART to combine the first 2.

Here are some details of the model,
but it's important to keep in mind that the reanalyses
look mostly like the actual weather,
while the specific CAM version has a somewhat marginal influence.
One point to note is that we use daily, quarter degree, sea surface temperatures from AVHRR
to force CAM.
I'm looking forward to Andrea's talk about the advantages of that SST resolution.

0:30



This reanalysis assimilated several million observations/day into the CAM model state, which we've defined by these variables.

We use wind and temperature observations from airplanes, radiosondes, and satellites,

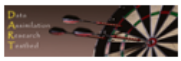
and GPS refractivity observations (basically, density).

0:20

What This Yields

80 equally likely CESM states consistent with
✓ the actual weather
✓ CAM6.0.34 physics.

"Consistent" = explicitly uses the 2 main sources of uncertainty
(observation errors and model ensemble spread)
to balance the information in the obs and in the model hindcast.



0:22

The Benefit of Lots of Observations

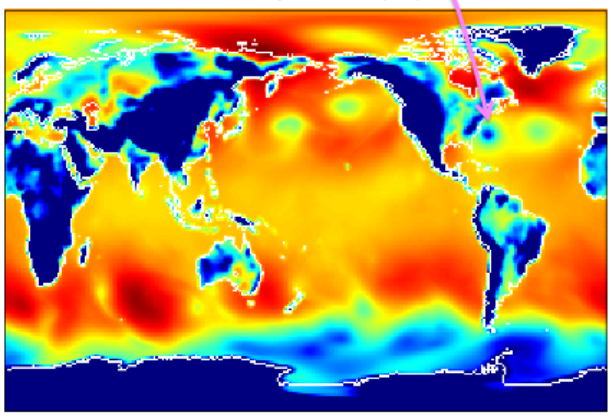
1 degree resolution model doesn't generate great hurricanes by itself,
but here's hurricane Sandy shortly before landfall 0Z 10/29/2012
in the analysis, ensemble mean, surface pressure.

Surface pressure (Pa)

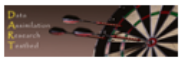
Central pressure:
NHC "best track" = 950 hPa
KCZ-PW* = 959 hPa
Ensemble mean = 956.6 hPa

Good, given:
- location +/- 60 km
- Ens. st. dev. = 1.7 hPa

*Knaff-Zehr-Courtney
pressure-wind relationship



Range; 950-1040 hPa, to highlight Sandy



CESM-ESPWG June 2020

6

This is good, give that
the location of the ensemble central pressure could be wrong by up to 60 km due to the 1 degree resolution,
and ensemble standard deviation of central pressures is 1.7 hPa,
so the ensemble is not inconsistent with the NHC numbers.
0:40

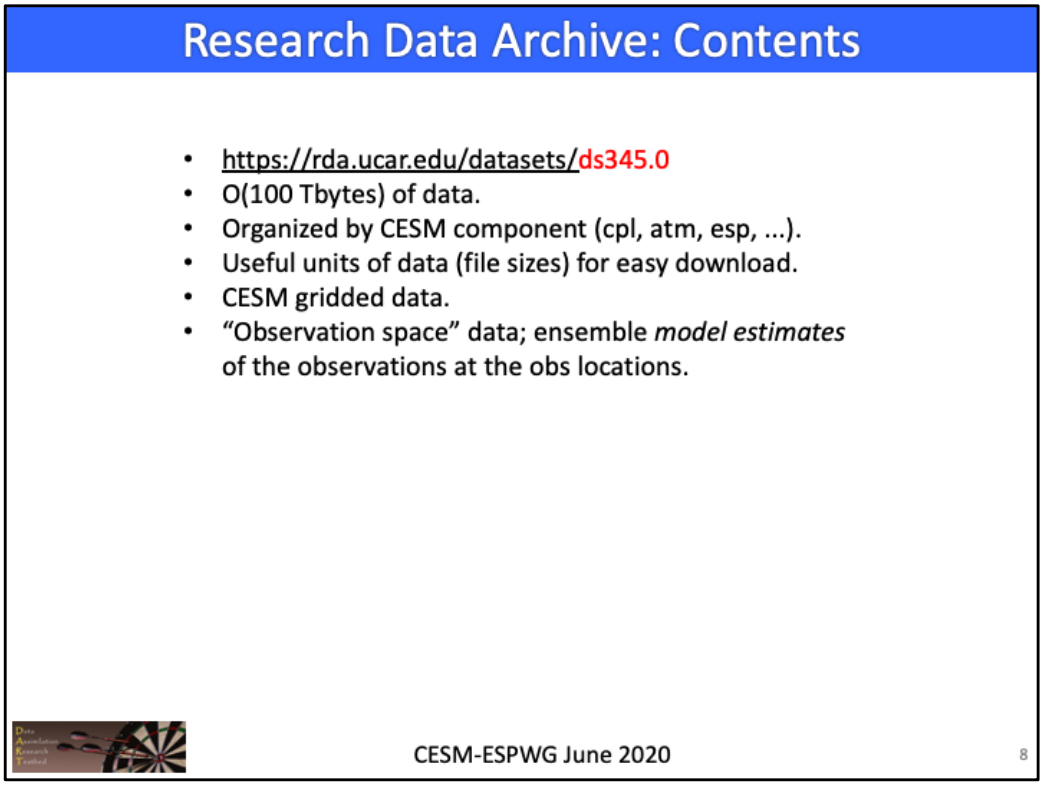
Unprecedented Data Set

Full ensemble output (LENS and previous ensemble DA have shown the value.)
4x per day (some variables) for 9 years.
A unique combination with many uses.

- ✓ Real world initial conditions for CAM (CLM, CICE, and MOSART?) ensembles with justifiable spread.
- ✓ Realistic atmospheric forcing of all surface components in simulations and assimilations, with justifiable variability.
- ✓ Model improvement through direct comparison with observations.
- ✓ Many theses and papers are waiting to be extracted.

Extensive, free data, which is largely unexplored.

0:50

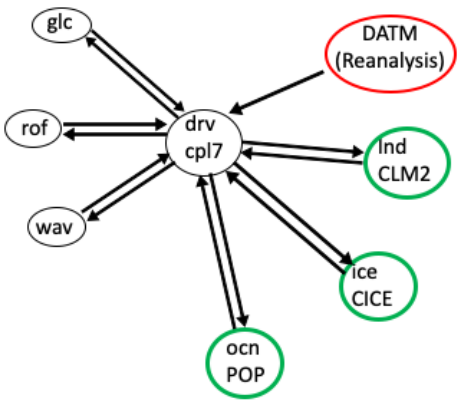


There's O(100) Tbytes of data archived in the (read)

Using much of this data requires no knowledge of data assimilation.
People interested in the DA will find a trove of useful and interesting data.

0:35

RDA 345.0: Atmospheric forcing of surface components



Cpl history files:

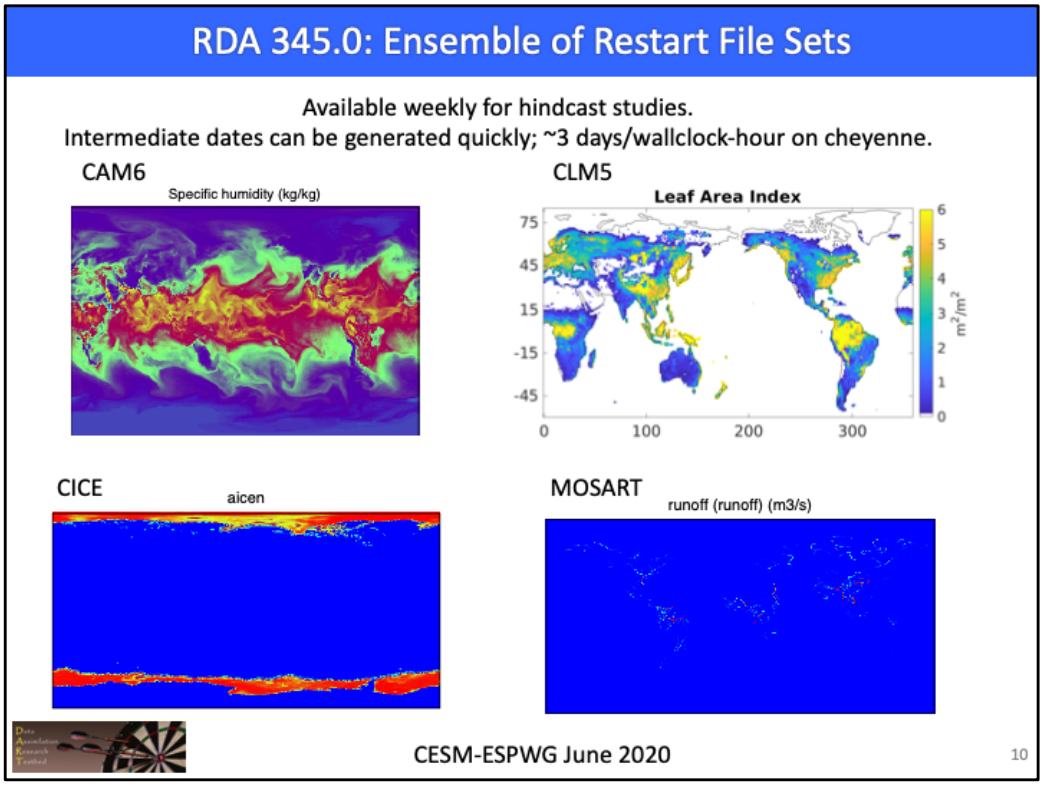
- frequencies ranging from 1-6 hours
- ready to use in CESM in DATM mode
- 1 year, 1 member per file.

2011-2019
Ensemble provides realistic variability.
Use for hindcasts of surface components:
CLM, POP(MOM), CICE, MOSART, CISM, WW3
including data assimilation experiments.

These models
have DART interfaces for assimilation.

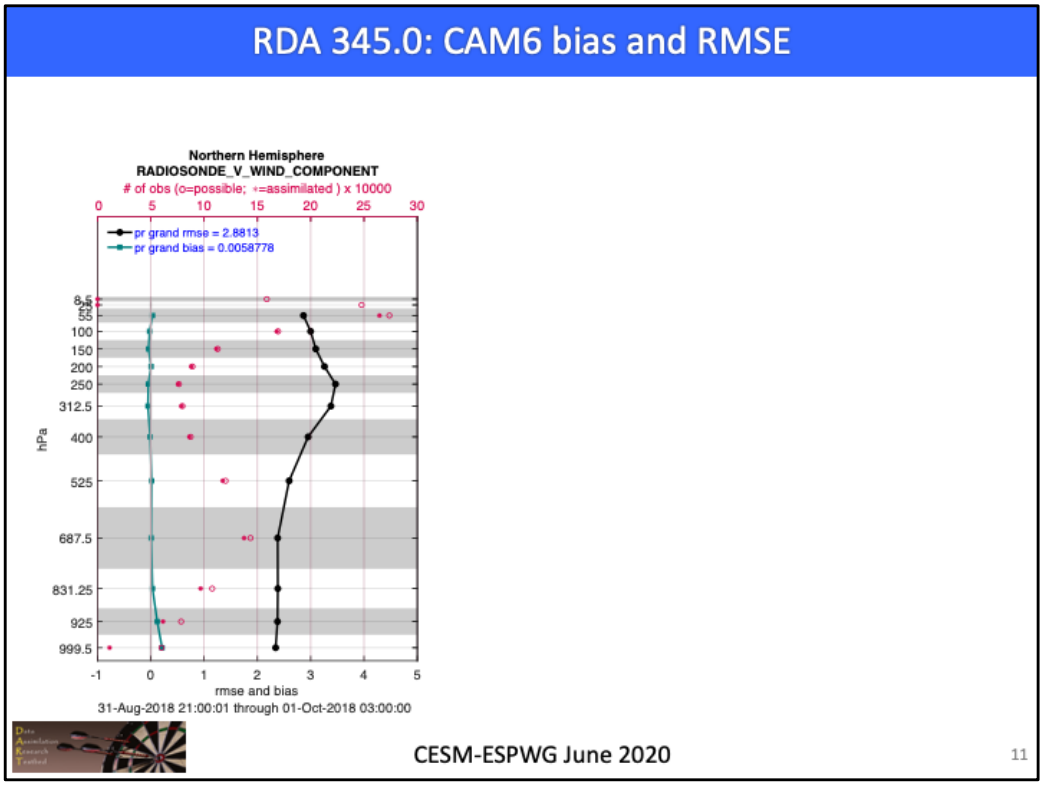
Here's an outline of how these products can be used.
The first is atmospheric forcing of surface components.
The forcing is contained in (read)

0:40



(Read)
Here's a snapshot from each of the 4 active components' restart files.
These are well-spun-up model states.

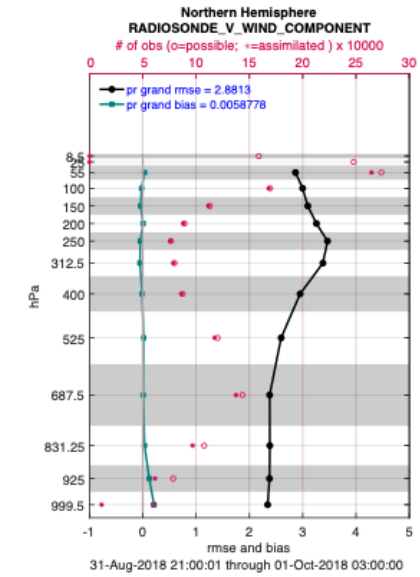
0:20



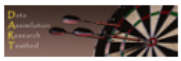
Other files can be used to identify ways that CAM6 struggles to recreate the weather.
Here are profiles of the bias and RMSE
relative to the radiosonde V wind component
in the northern hemisphere.
Not much room for improvement in the bias!
It also shows the number of observations available, and the number used.
This tells us how much confidence to have in the curves;
lots of observations used = higher confidence.

0:30

RDA 345.0: CAM6 bias and RMSE



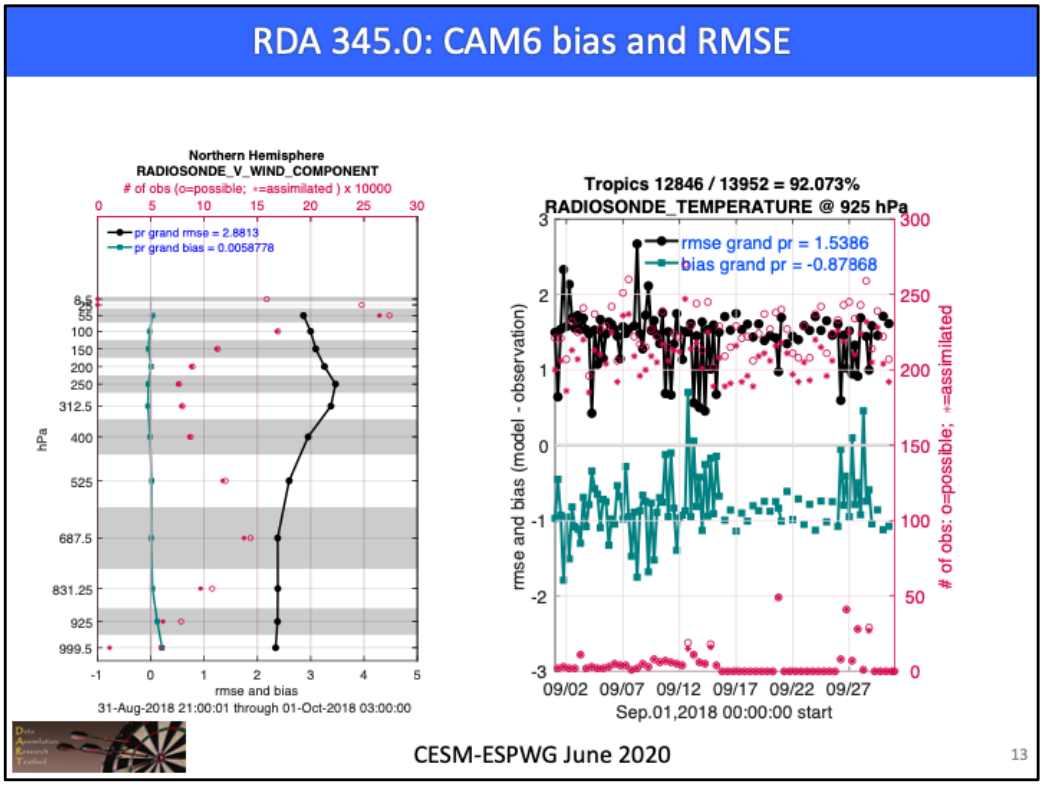
They're not all this good!



CESM-ESPWG June 2020

12

0:05



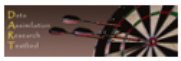
Here's a time series plot of the bias and RMSE
relative to the radiosonde temperature observations
in the Tropics in a layer around 925 hPa.
The numbers of observations are still in red.
This shows a persistent T bias of almost 1 K,
which is also apparent relative to all the other observation platforms
in all seasons.

0:25

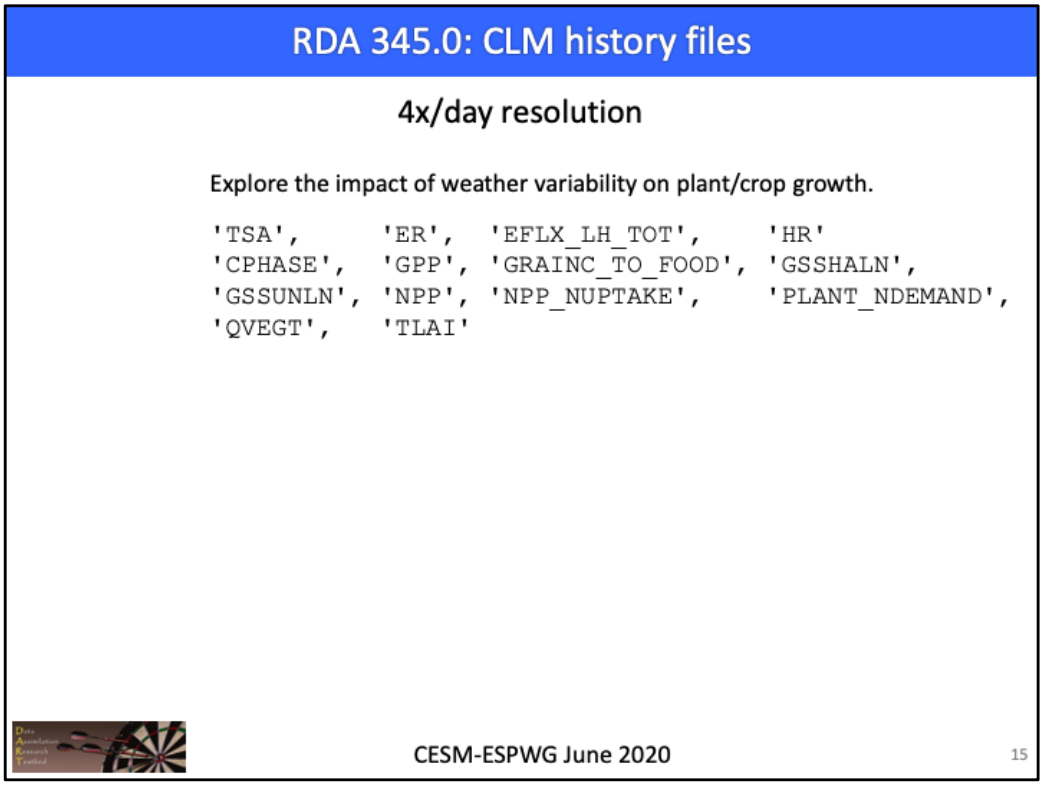
RDA 345.0: Assimilation Increments

Not comfortable in observation space?

On CAM's grid see the spatial and temporal distribution
of the corrections to the model state
caused by the observations.
Ensemble mean available 4x/day, members weekly.



0:18



We've written out some fields to CLM history files,
to (read Explore...)

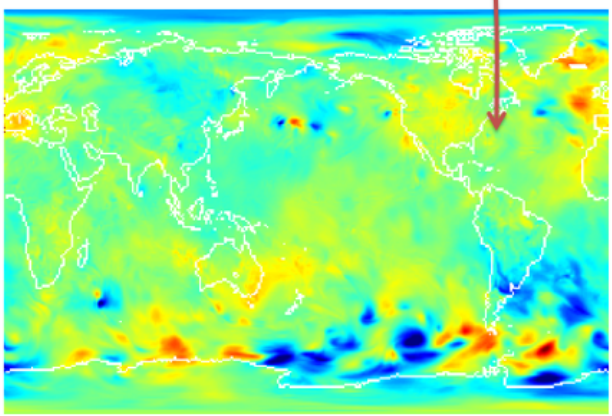
0:12

RDA 345.0: Ensemble Member 1 Difference from Mean

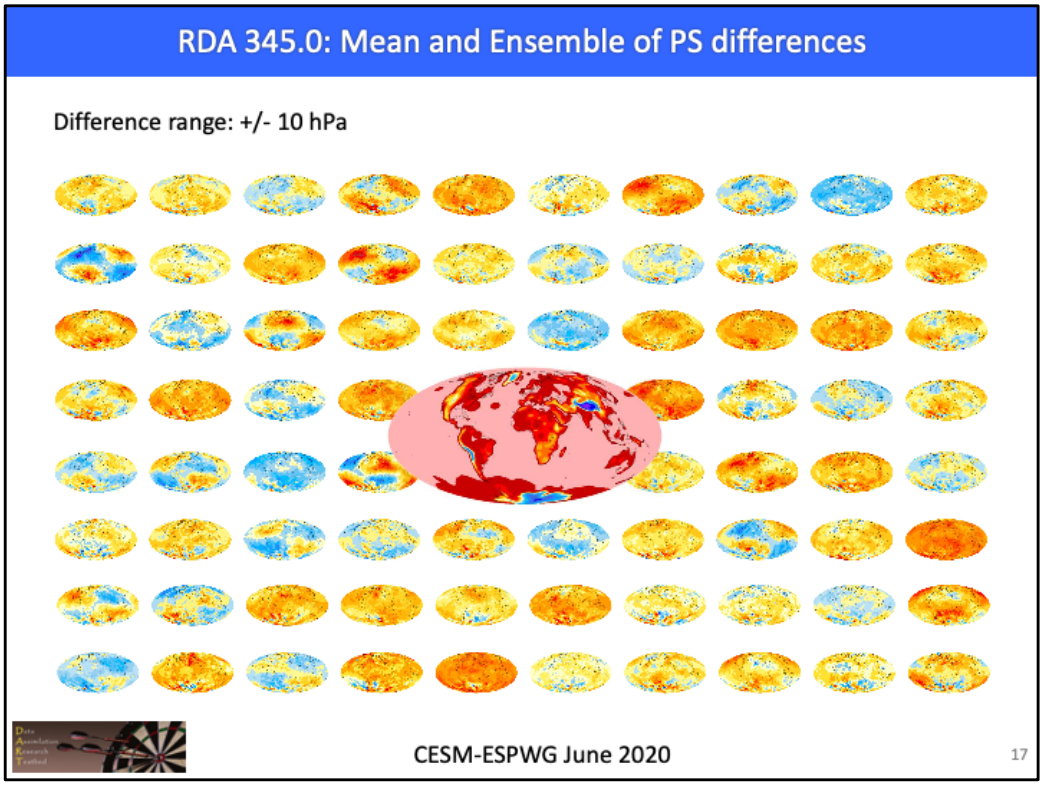
Sandy in member 1 is very close to Sandy in the mean.

Large differences; where there are few obs. and/or large error growth.

central pressure:
mean 956.6
mem1 957.1

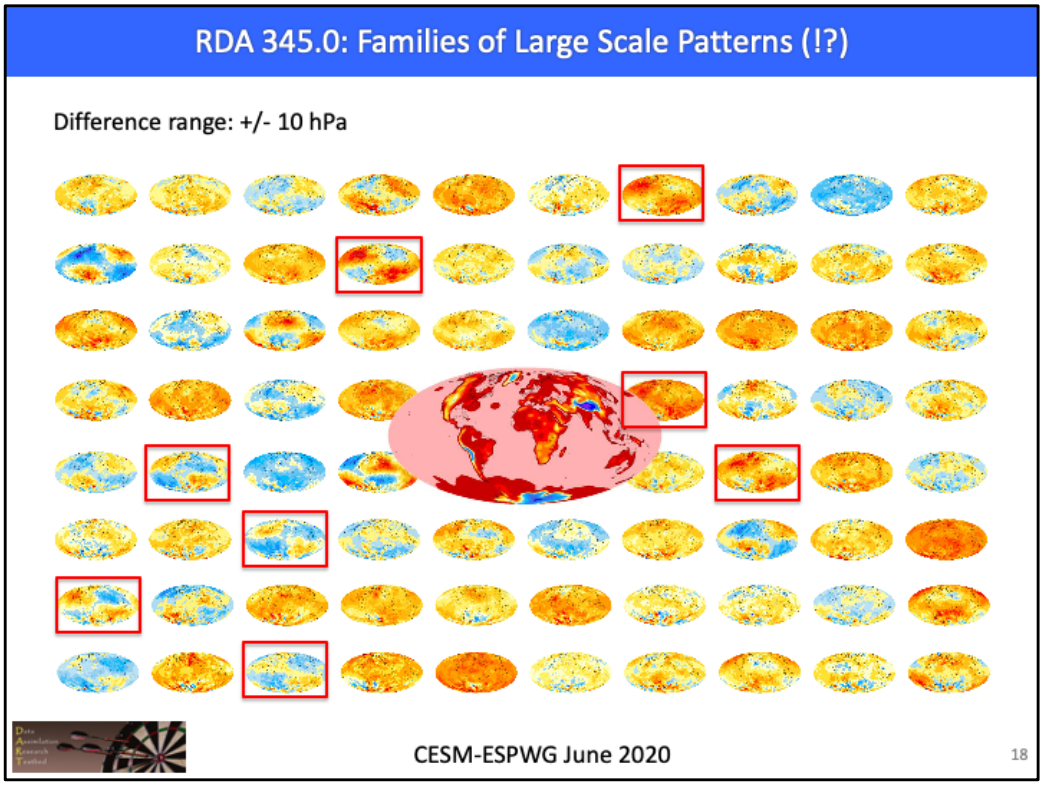


0:30



Here's the mean and the whole ensemble of PS differences from the mean viewed from a million miles high.
There's a healthy variety of differences,
all equally likely and consistent with available obs.
But your eye may already be picking out members that seem very similar.

0:20



We have no idea what determines these large scale patterns,
which is an interesting DA question, and could possibly inform us about the model.

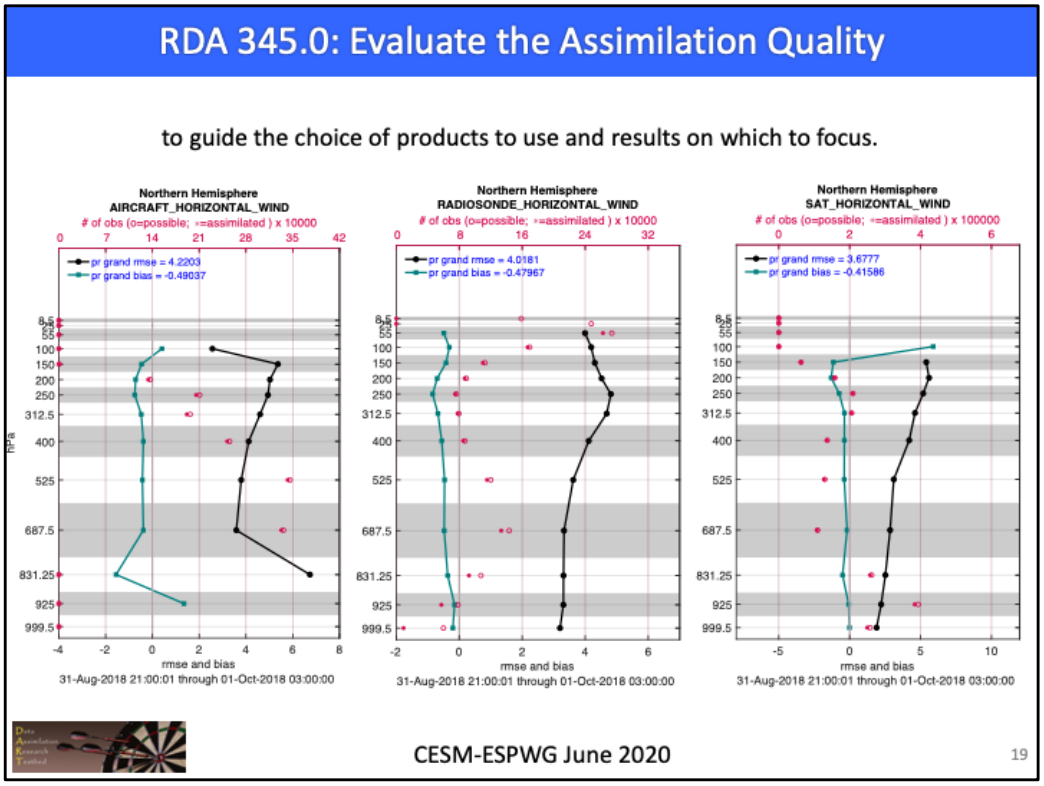
Are they meaningful?
Are there distinct groups?
Is it just a brain finding patterns in random data?

It could be a great project to figure this out:

semester or summer?
master's?
Ph.D. if done the right way?

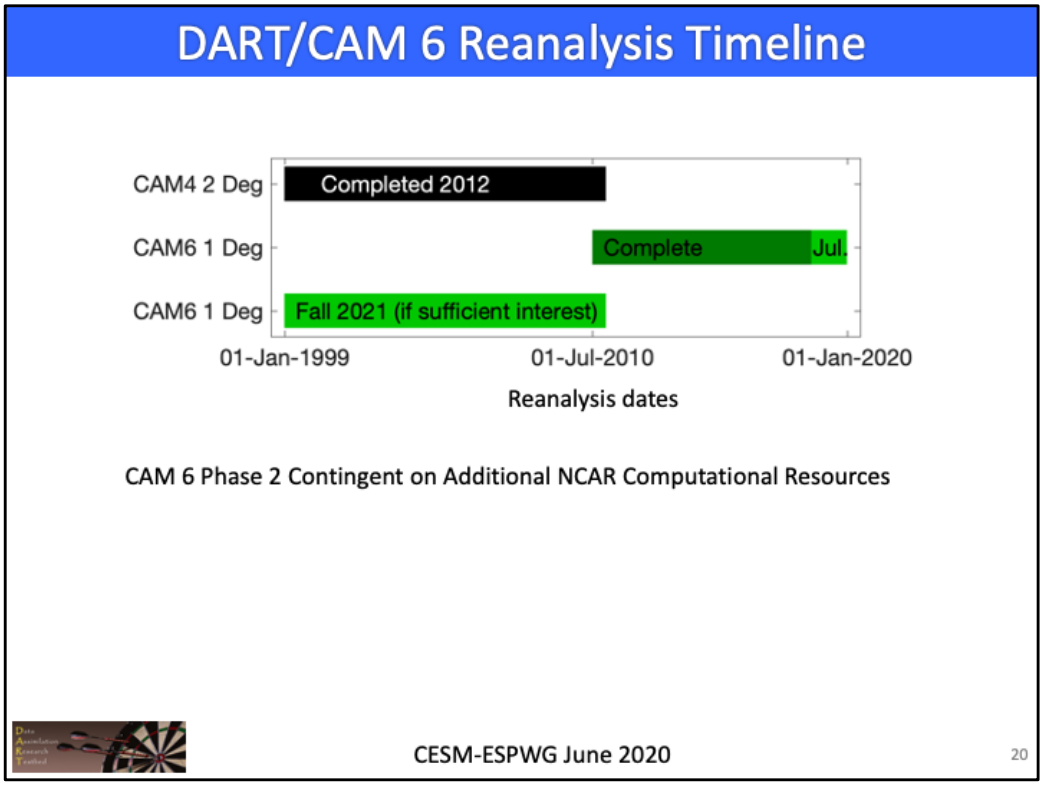
8 years of data (and counting) surely have more surprises.

0:30



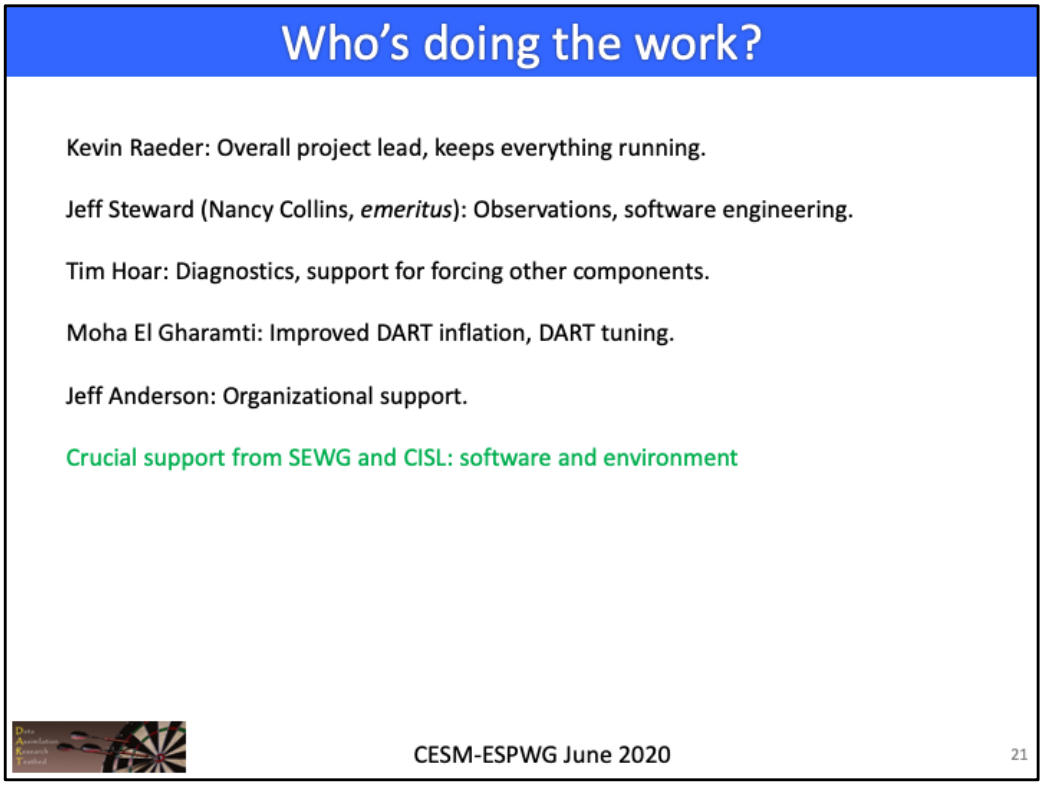
There are ways to evaluate the assimilation quality
to guide the choice of products to use and results on which to focus.
Here I'm showing the bias and RMSE relative to horizontal wind
from several sources.
(left figure)
In the middle layers things look good,
but in the lower layers, maybe not.
Then we should look at the number of observations used;
hardly any in the lower layers, so the reported bias is not reliable.
But the biases relative to other wind observations look even better
in the lower layers, and they have lots of observations.

0:35



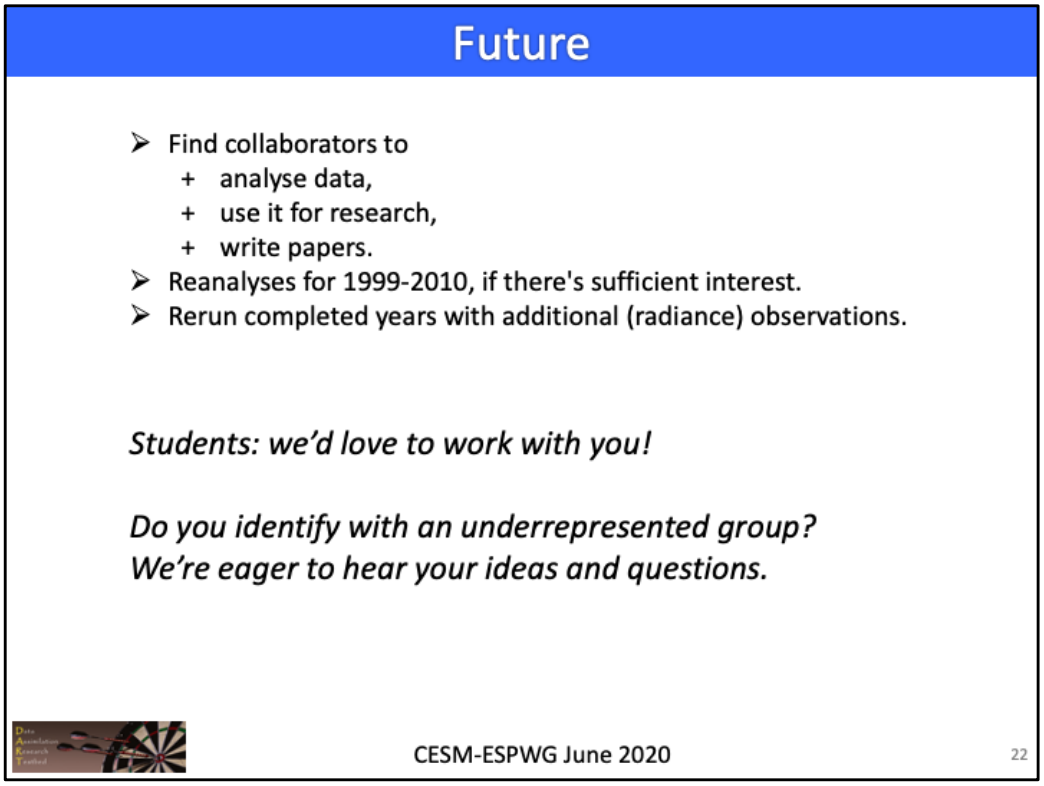
Here's the status of the project.
The first reanalysis, using the 2 degree CAM4, was completed 8 years ago.
The current, 1 degree CAM6 reanalysis has years 2011-2018 completed,
with 1 more year to go.
That should be finished in the next few weeks.
We may run the reanalysis for 1999-2010,
if there's enough community interest and computer time.

0:30



Here's a list of who's doing the work

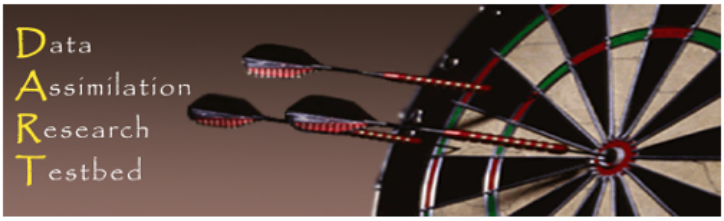
0:05



Our top priority is to (read)

0:20

For more information:

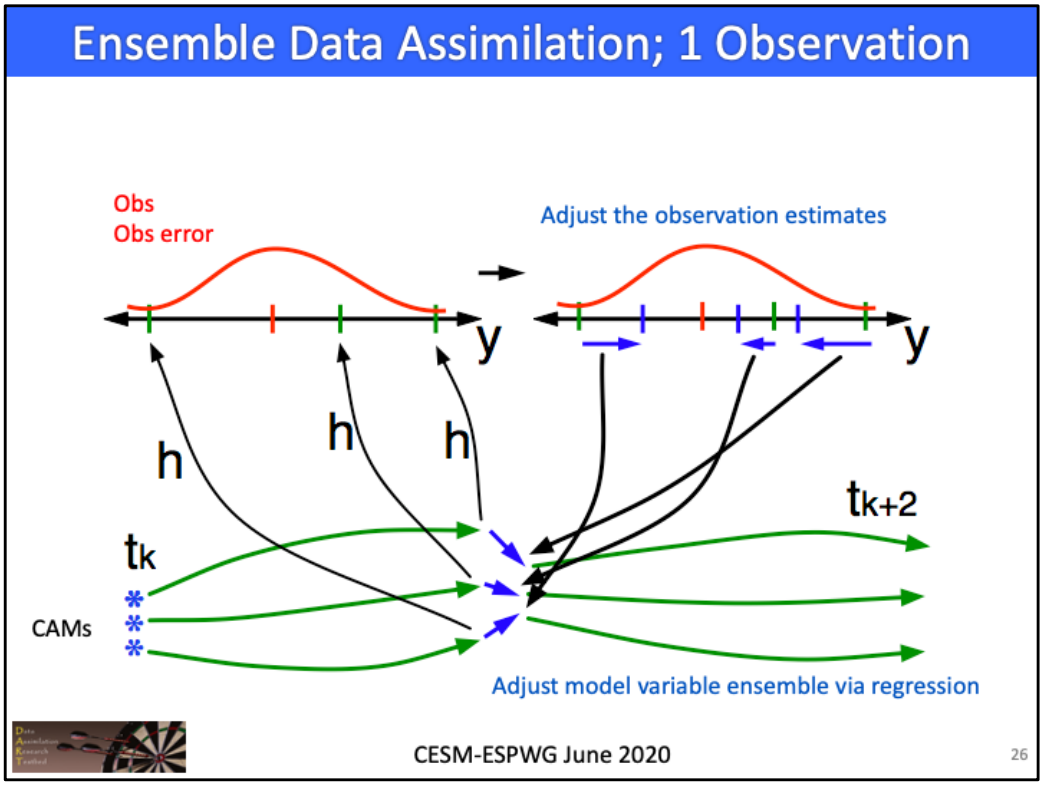


https://dart.ucar.edu

dart@ucar.edu

CISL Computational & Information Systems Lab

We would 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.



Extremely brief review and "I'm happy to walk you through this later, or see the DART tutorial"

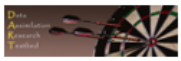
Data assimilation is essentially guiding a model state
to be consistent with the information in available observations.

We start with an ensemble of differing model states.
We run the model forward to the time of an observation.
We use each member to calculate an estimate of the observation.
The observation is often not a model state variable.
We use statistics to adjust the estimated observations towards the actual observation.
Then we use linear regression to determine
the correlation between the observations and a model state variable
and to adjust the model state variable accordingly.
The adjusted state variables are called the "analysis"
0:35

Reanalysis Quick Facts: Assimilation

Assimilation:

- Ensemble Adjustment Kalman Filter
- 6-hour window around each observation time.
- Enhanced adaptive inflation (Gharamti 2017).
- Tuned parameters for localization, inflation.
- DART Manhattan.



CESM-ESPWG June 2020 27

I am interested in Ensemble DA which involves running multiple instances of the model that differ only in their initial state – this represents our uncertainty in our knowledge of the state of the system.
The first public release of DART was in 2004 and the seminal review paper is from 2009.

~~I want to make sure that we understand each other and when I say 'data assimilation' I am referring to confronting a model state with observations and updating the model **state** to reflect the information in the observations. DART additionally allows one to augment the state with parameters which can be estimated – which is of interest to the rest of the team – but that's a topic for a different talk.~~

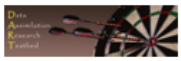
Reanalysis Quick Facts: Observations

Observations assimilated:

- Temperatures and winds from radiosondes, ACARS and aircraft.
- Cloud motion vector winds.
- GPS radio occultation refractivity.
- AIRS temperature retrievals.

Observations evaluated:

- Radiosonde specific humidity.
- AIRS specific humidity retrievals.
- Radiosonde, land and marine altimeter.

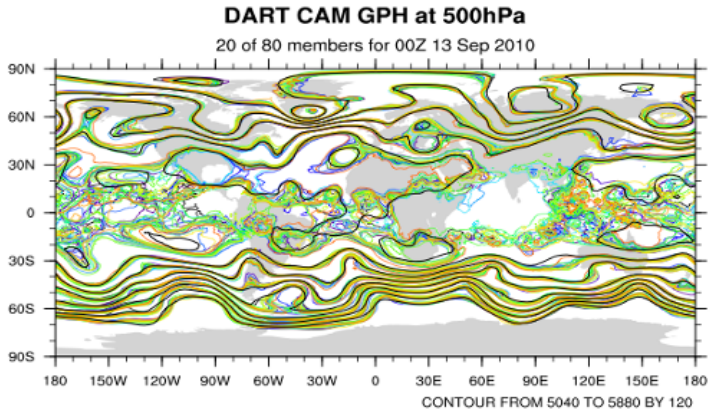


CESM-ESPWG June 2020 28

I am interested in Ensemble DA which involves running multiple instances of the model that differ only in their initial state – this represents our uncertainty in our knowledge of the state of the system.
The first public release of DART was in 2004 and the seminal review paper is from 2009.

~~I want to make sure that we understand each other and when I say 'data assimilation' I am referring to confronting a model state with observations and updating the model **state** to reflect the information in the observations. DART additionally allows one to augment the state with parameters which can be estimated – which is of interest to the rest of the team – but that's a topic for a different talk.~~

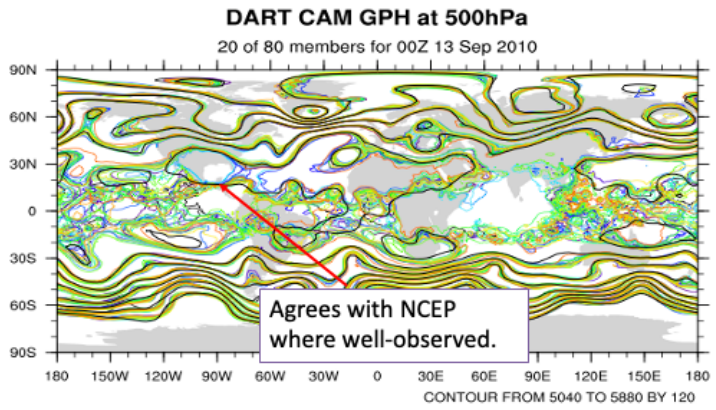
An Ensemble Reanalysis with CAM in CESM: Results

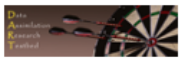


Color contours from DART (20 of 80 ensemble members). Show Uncertainty.

Black from operational NCEP FNL analysis.

An Ensemble Reanalysis with CAM in CESM: Results



Color contours from DART (20 of 80 ensemble members). Show Uncertainty.

Black from operational NCEP FNL analysis.