DART and CESM

Helen Kershaw
IMAGe/DARes

Nancy Collins, Jeff Anderson, Tim Hoar, Kevin Raeder
Four Topics

A little about DART

DART and CESM so far

What’s next for DART-CESM

Preparation for what’s next
A little about DART
A little about DART

Data Assimilation Research Testbed
Ensemble Data Assimilation
Group of model forecasts
Ensemble Data Assimilation

Group of model forecasts

Measurements
Ensemble Data Assimilation

Group of model forecasts

Measurements

Improved estimate
Ensemble Data Assimilation

CESM multi-instance

Group of model forecasts

Improved estimate

Measurements
The State
The State

pressure
temperature
vapor mixing ratio
The State

DART state vector

pressure
temperature
vapor mixing ratio
The State

- pressure
- temperature
- vapor mixing ratio

multiple copies
The State

- pressure
- temperature
- vapor mixing ratio

multiple copies
The State

pressure
temperature
vapor mixing ratio

multiple copies
• Apply observation updates in parallel
• Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
• Apply observation updates in parallel
• Round robin layout for load balancing
  - localization
Assimilation

- Applying observation updates in parallel
- Round robin layout for load balancing
  - Localization
Assimilation

• Apply observation updates in parallel
• Round robin layout for load balancing
  - localization
• Apply observation updates in parallel
• Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
• Apply observation updates in parallel
• Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Assimilation

- Apply observation updates in parallel
- Round robin layout for load balancing
  - localization
Data storage

Whole model state available to a single processor

All copies of some variables available to a single processor
Data storage

Whole model state available to a single processor

All copies of some variables available to a single processor
Data storage

Whole model state available to a single processor

All copies of some variables available to a single processor
DART and CESM

so far ...
DART Multiple Component Data Assimilation (Multiple Active Components)

DART assimilates observations into components separately

Coupler moves the components to the next time step
DART-CESM

**Single-Component Assimilations**

Yongfei Zhang *(University of Texas)*:
  Assimilating snow cover fraction into the land model

Nick Pedatella *(UCAR COSMIC)*:
  WACCM (Whole Atmosphere Community Climate Model) assimilating TIMED/SABER and aura MLS satellite measurements

Andrew Fox *(NEON)*:
  Assimilating tower fluxes of latent heat (LE), sensible heat (H), and net ecosystem production into the land model
DART-CESM

Multi-Component Assimilations

Abhishek Chatterjee *(NCAR)*:
  coupled atmosphere-ocean-land

Alicia Karspeck *(NCAR)*:
  coupled atmosphere-ocean and ocean
What’s Next?
What’s Next?

Cross-component data assimilation
Cross-Component Data Assimilation
Cross-Component Data Assimilation
Cross-Component Data Assimilation

COUPLER

CAM

CLM

ROF

POP

CICE

observations

DART
Cross-Component Data Assimilation

Observations impact all components
Preparing for Cross-Component DA
Preparing for Cross-Component DA

node

node

CESM state vector

CESM state vector

CESM state vector

CESM state vector
Preparing for Cross-Component DA

- node
- node
- CESM state vector
- CESM state vector
Preparing for Cross-Component DA

- CESM state vector
- node
Why do we have the whole state?
Why do we have the whole state?

For each state element:

  Where am I?
  Which observations are nearby?
Why do we have the whole state?

For each observation:

Where am I?
Which observations are nearby?
What does each state think I should be?
Calculation of the forward operator

Or, what a state thinks the observation should be
Calculation of the forward operator

Or, what a state thinks the observation should be
Limitations caused by data storage

- Hard minimum on calculation time
- Hard maximum on model size
- You have to move all your data
What do we do?

- Hard minimum on calculation time
- Hard maximum on model size
- You have to move all your data
Don’t move the data
Don’t move the data

Use **one sided communication** to grab state elements when needed
Don’t move the data

Use **one sided communication** to grab state elements when needed
Don’t move the data

Use **one sided communication** to grab state elements when needed

Reduce data movement
Don’t move the data

Use **one sided communication** to grab state elements when needed

Reduce data movement

Removes hard memory limit
Don’t move the data

Use **one sided communication** to grab state elements when needed

Reduce data movement

Removes hard memory limit

Vectorization of forward operator calculations
One-sided communication
One-sided communication

Me
One-sided communication

Me
One-sided communication

Me
One-sided communication

Me

Everyone Else
One-sided communication

Me

Everyone Else
One-sided communication

Me

window

Everyone Else
One-sided communication

Me

window

Everyone Else
One-sided communication

Me

Everyone Else

window
One-sided communication
One-sided communication
One-sided communication

1

Me

2

3
One-sided communication

Me

1

2

3
One-sided communication

Me

1 2 3
One-sided communication

Me

1

2

3
One-sided communication
Toy 10,000 element model

500,000 observations

Time

Cost
CAM FV forward operator
Specific humidity only: 23 090 observations

<table>
<thead>
<tr>
<th>processors</th>
<th>512</th>
<th>4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>old</td>
<td>1.01s</td>
<td>0.96s</td>
</tr>
<tr>
<td>new</td>
<td>0.73s</td>
<td>0.18s</td>
</tr>
</tbody>
</table>

CONTOUR FROM 5200 TO 5700 BY 100
More scalable forward operator

Memory
More scalable forward operator

Memory
More scalable forward operator

Memory
More scalable forward operator

Memory

Calculation

4 tasks doing all observations for 1 copy
More scalable forward operator

Memory

Calculation

4 tasks doing all observations for 1 copy

Lots of tasks doing some observations for all copies
Further Complications

Or, software engineering concerns
Further Complications

Or, software engineering concerns

Need to remain user extensible
Questions?

http://www.image.ucar.edu/DARes/DART

dart@ucar.edu
Further Complications
Or, software engineering concerns

What about all the users who are happy with DART as it is?
Further Complications

Or, software engineering concerns

What about all the users who are happy with DART as it is?
IO
Models do not run ensemble complete
Models do not run ensemble complete
Models do not run ensemble complete
Models do not run ensemble complete
Models do not run ensemble complete

You have to move data from the model to DART
Ideally:
Ideally:

Never looks like this in memory
All DART requires is that there are multiple model forecasts
Multi-instance forecasts create the ensemble
Multi-instance forecasts create the ensemble

- Model run ~10000 tasks
- Restart files for each model

IO

ensemble members

time

CESM
Multi-instance forecasts create the ensemble
Multi-instance forecasts create the ensemble
But vectorization is not perfect:

An observation can be in different model levels depending on the state