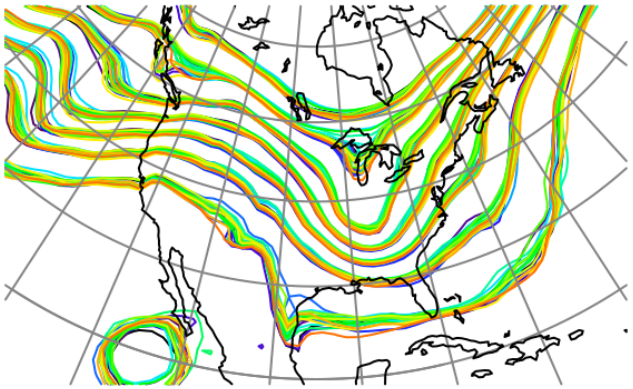
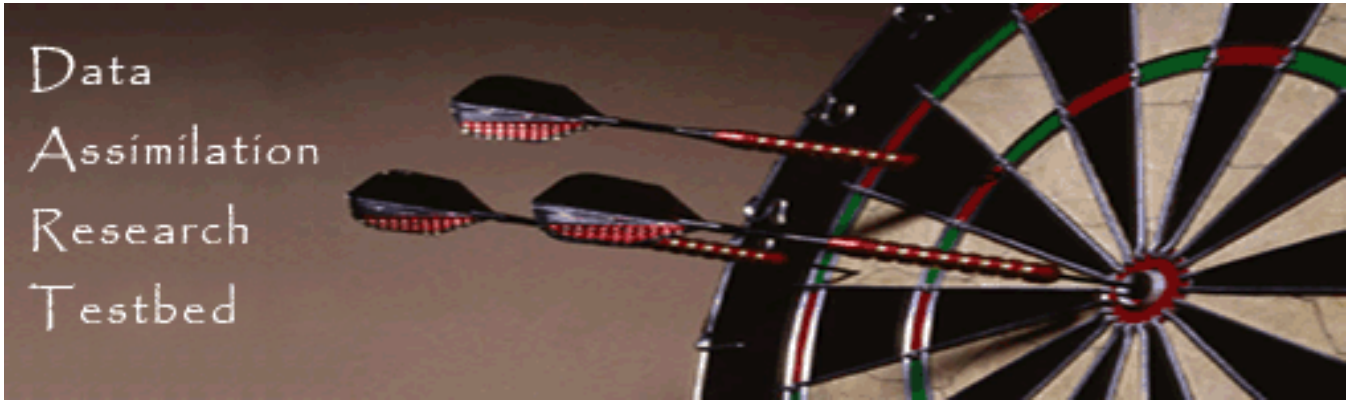


Scalable Computing Challenges in Ensemble Data Assimilation

Nancy Collins representing the NCAR Data Assimilation Research Section

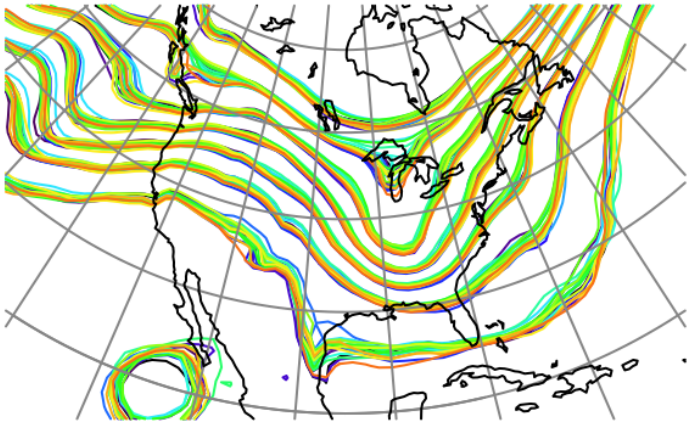


Presentation Outline

- Quick overview of Ensemble Data Assimilation
- Overview of the DART System
- Discussion of supported models and observations
- Current work for new highly parallel systems

What is Data Assimilation?

Observations combined with a Model forecast…



…to produce an analysis
(best possible estimate).

Data Assimilation Types

- Variational Systems
  - Used by operational NWP forecasting centers
- Ensemble Systems
  - Make many forecasts and use statistical methods
  - Easier to develop a DA system, especially for large models
  - Feasible for individual researchers, small groups
  - Produces uncertainty information

Ensemble Filter for Large Geophysical Models

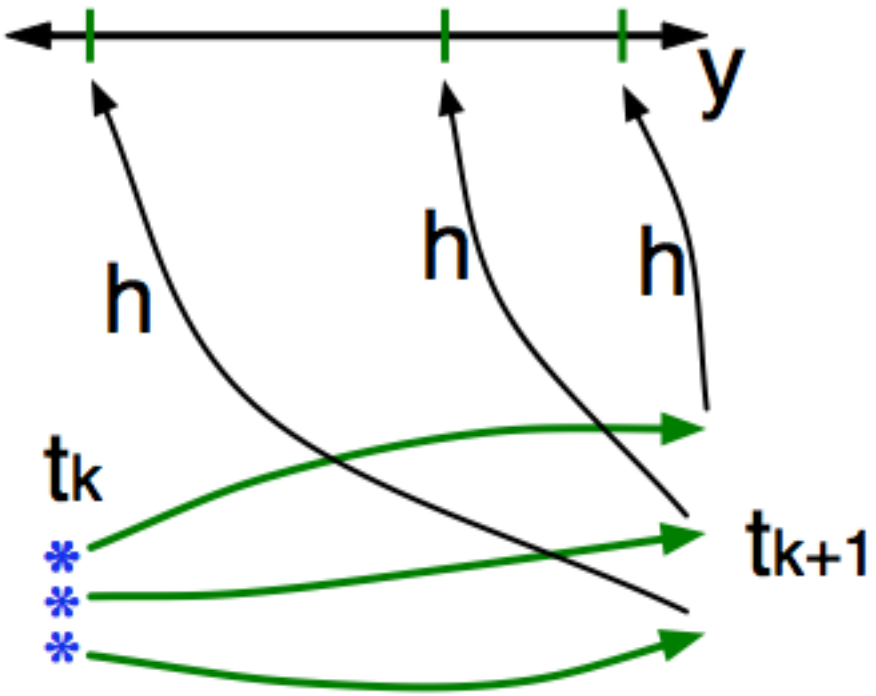
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Filter for Large Geophysical Models

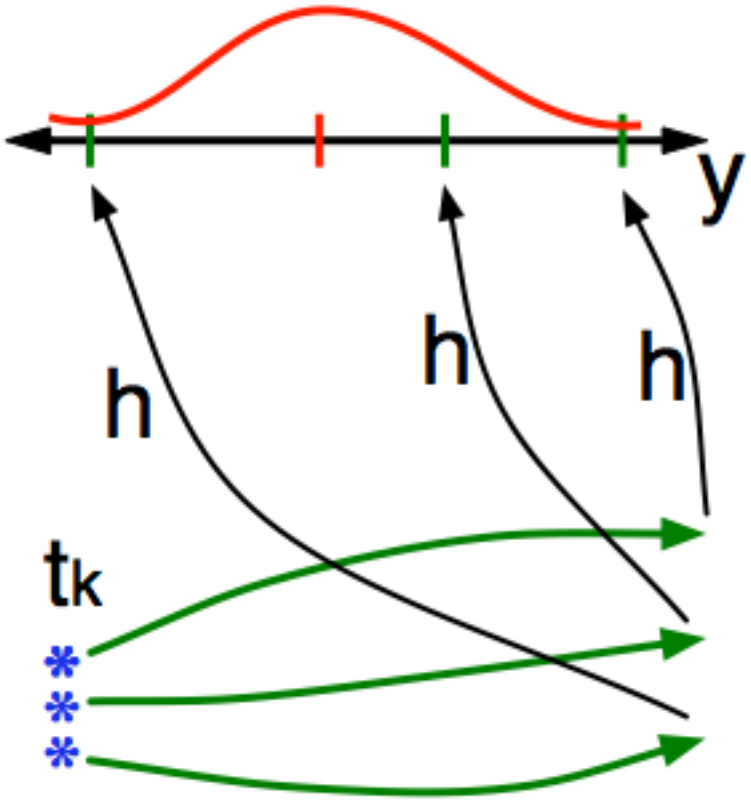
2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.



Theory: observations from instruments with uncorrelated errors can be done sequentially.

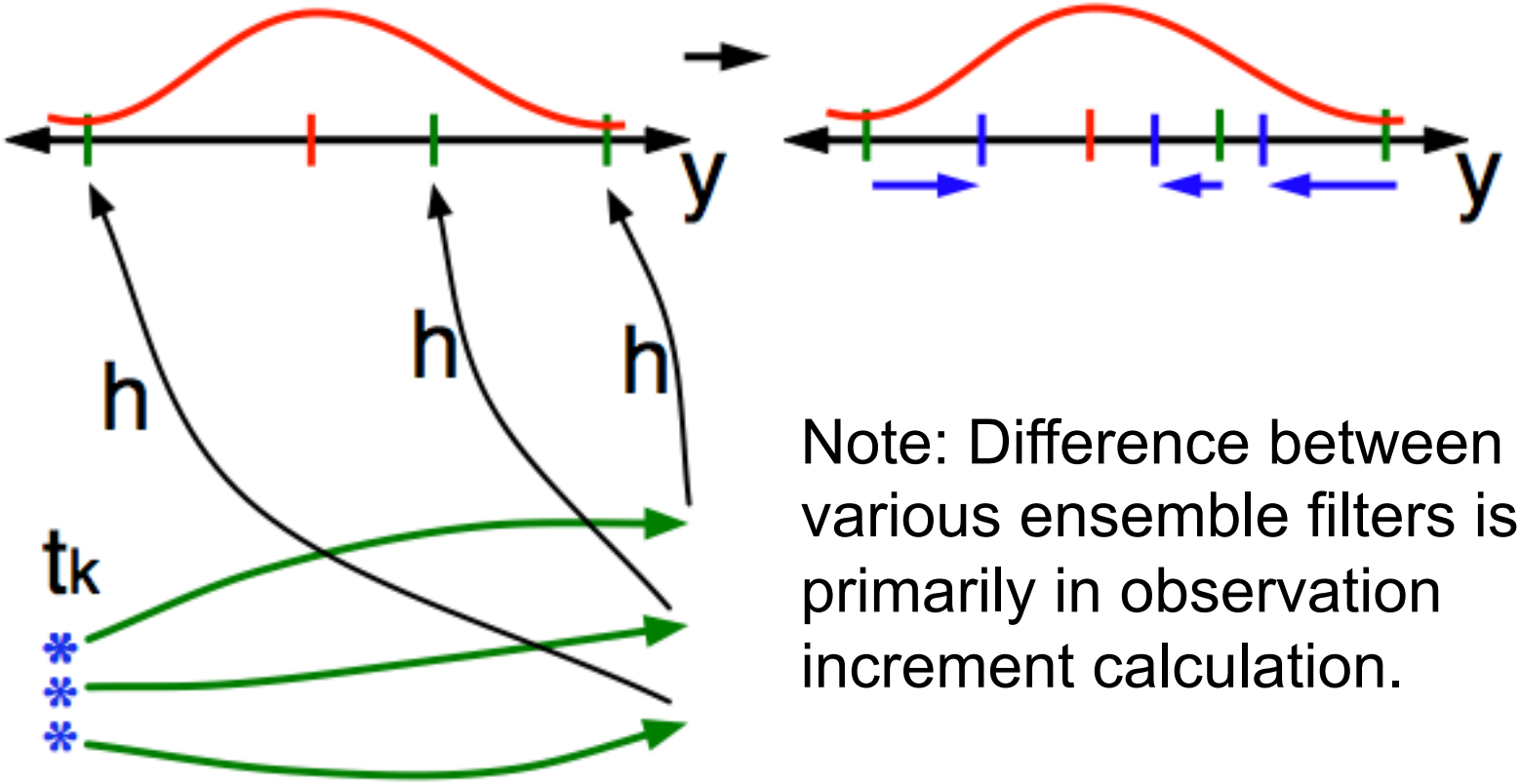
Ensemble Filter for Large Geophysical Models

3. Get observed value and observational error distribution from observing system.



Ensemble Filter for Large Geophysical Models

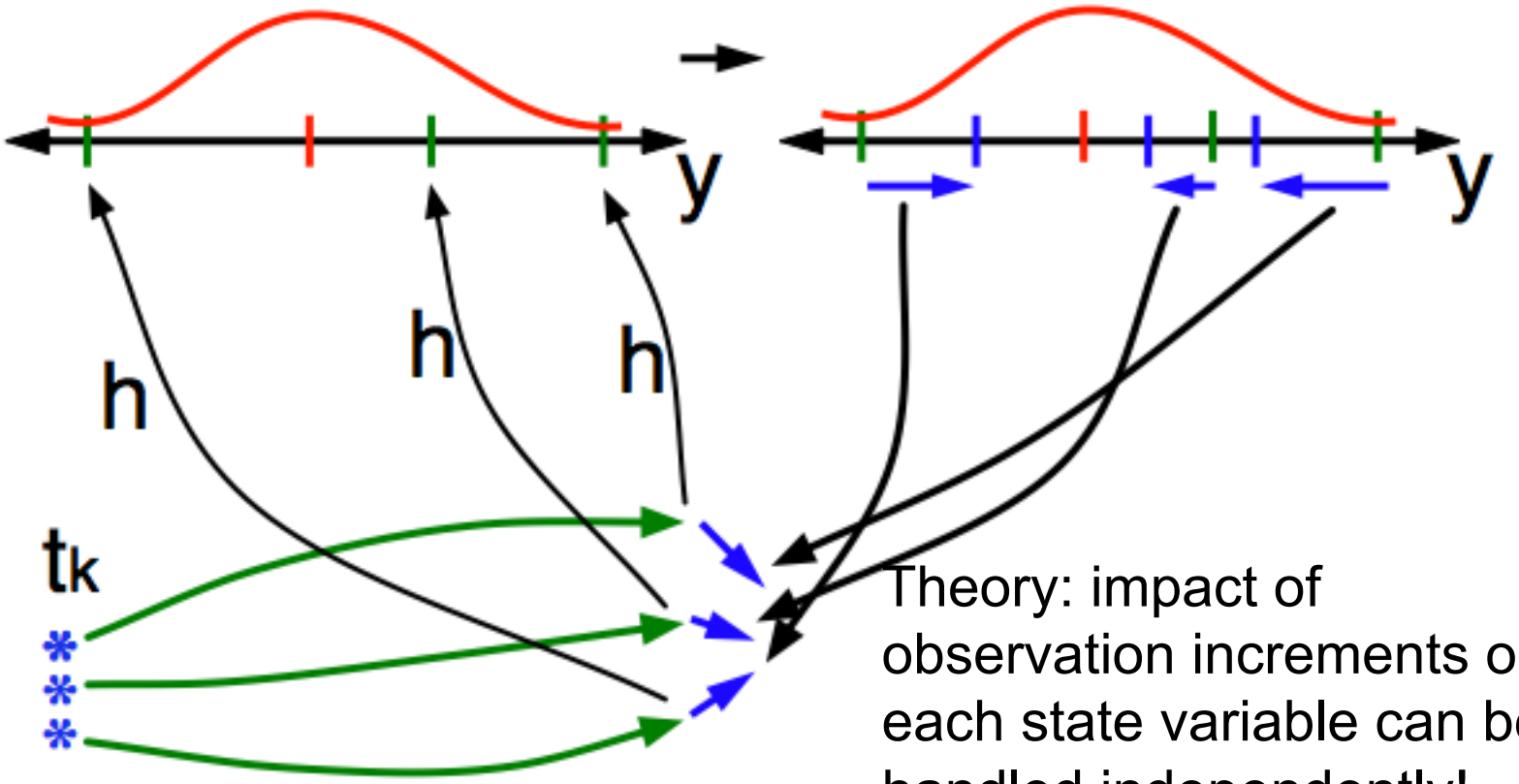
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



Note: Difference between various ensemble filters is primarily in observation increment calculation.

Ensemble Filter for Large Geophysical Models

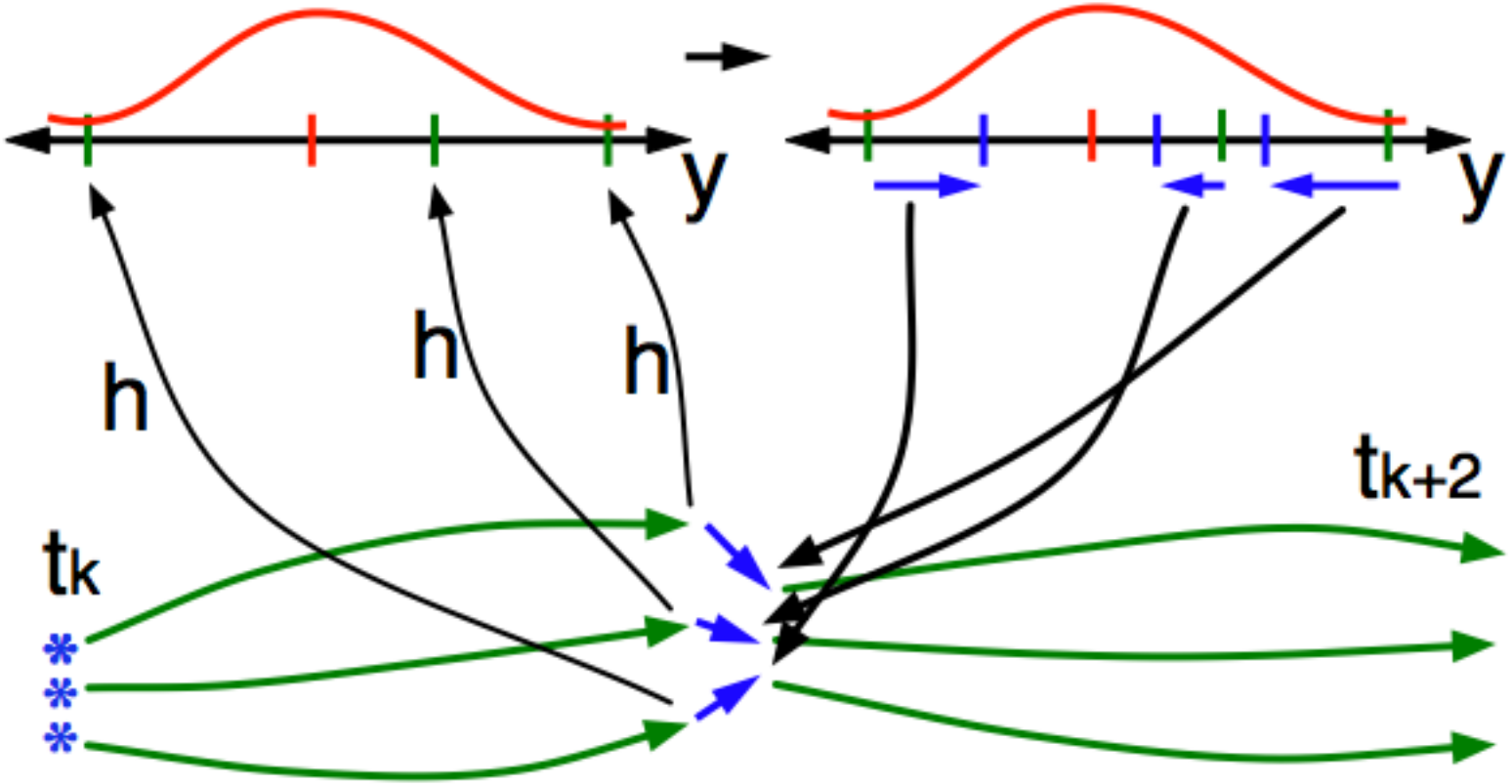
5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



Theory: impact of observation increments on each state variable can be handled independently!

Ensemble Filter for Large Geophysical Models

6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …



Data Assimilation Research Testbed

- A state-of-the-art Data Assimilation System for Geoscience
  - Flexible, portable, well-tested, extensible, free!
  - Works with many models.
  - Works with any observations: Real, synthetic, novel.
- A Data Assimilation Research System
  - Theory based, widely applicable general techniques.
  - Localization, Sampling Error Correction, Adaptive Inflation, …
- A Teaching Tool
  - Extensive tutorial materials with examples, exercises, explanations.
- People: The DAReS Team

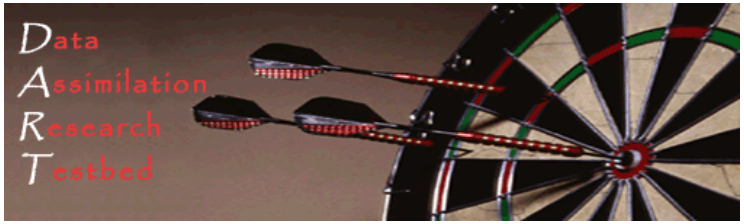
Running DART

- Users can run DART:
  - With no code changes
  - Add their own new models
  - Add their own new observation types
  - Change the assimilation algorithms

- DART can run on:
  - Supercomputers, Clusters with MPI parallelism
  - Workstations, Laptops with or without MPI

DA Techniques Available to All Models

- Localization
  - Restrict the search area and impact of observations to spatially close locations
- Adaptive Inflation
  - Automatically increase spread where needed without altering ensemble mean to maintain ensemble spread
- Sampling Error Correction
  - Improve performance of smaller ensemble sizes

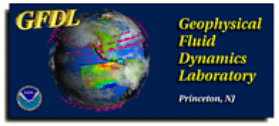


DART is used at:

48 UCAR member universities,
More than 100 other sites,
(More than 1500 registered users).

UNIVERSITY OF WASHINGTON

CIRES



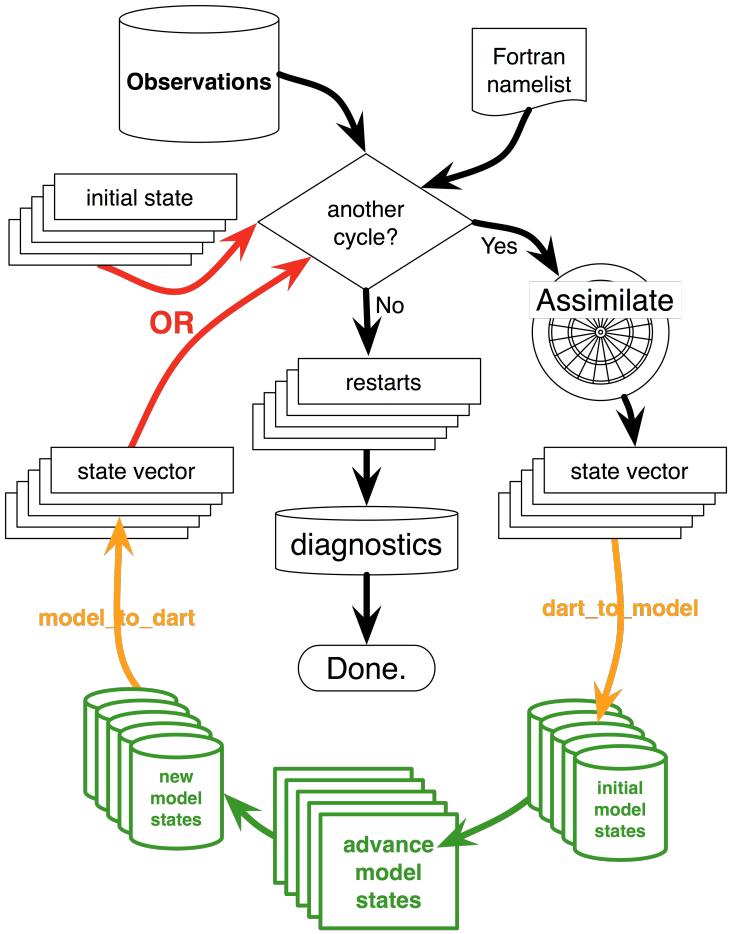
THE UNIVERSITY OF UTAH

PENNSTATE

JPL
Jet Propulsion Laboratory
California Institute of Technology

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

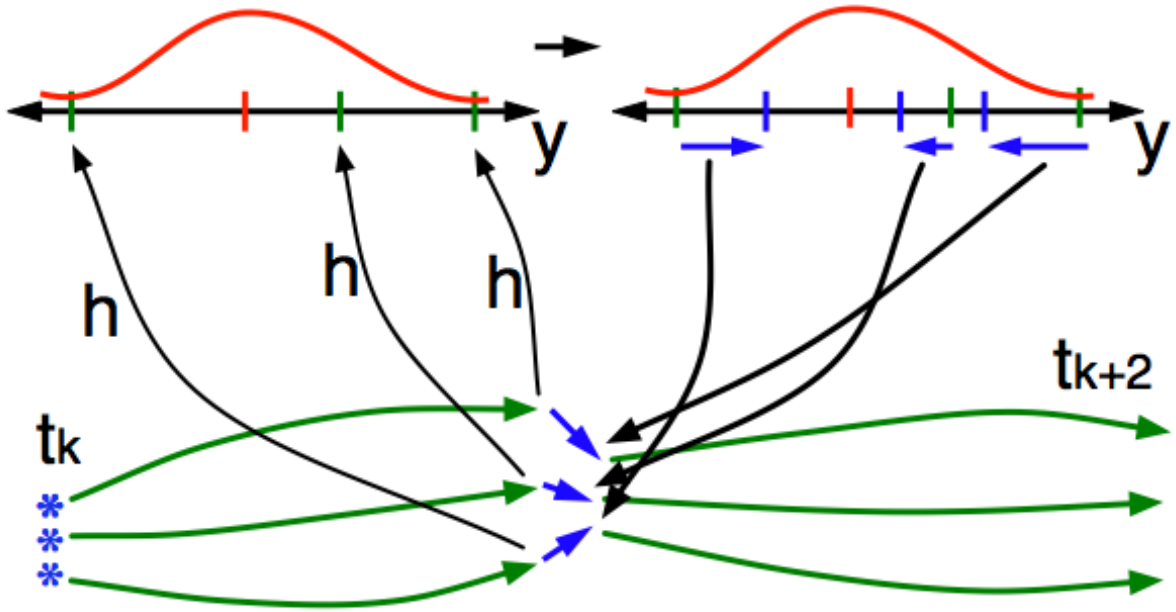
How DART runs with Large Models



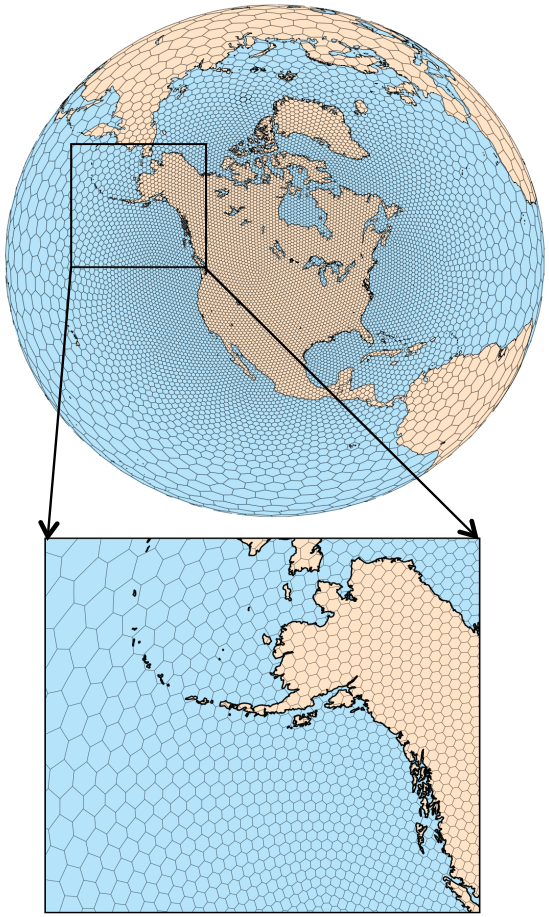
Adding New Models to DART

To work with a new model, DART needs:

1. A way to make model forecasts (done by modelers)
2. Transfer model data to/from a 1D vector (modelers)
3. Forward operators ‘h’ - Interpolation (collaboration)



Unstructured Grids

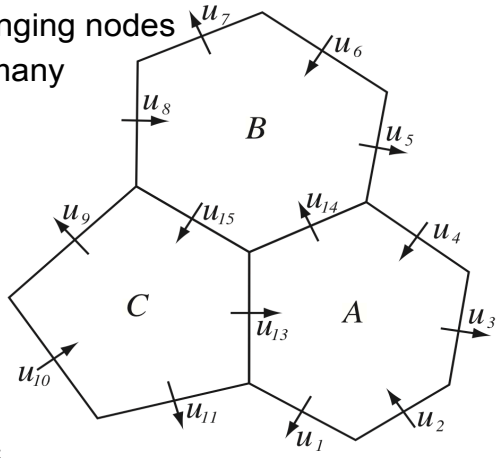


Spherical Centroidal Voronoi Tessellations (SCVT)

- Mostly *hexagons*, some pentagons and 7-sided cells.
- Lines connecting cell centers intersect cell edges at right angles.
- Lines connecting cell centers are bisected by cell edge.
- Uniform resolution – traditional icosahedral (hexagonal) mesh.

Conformal grid

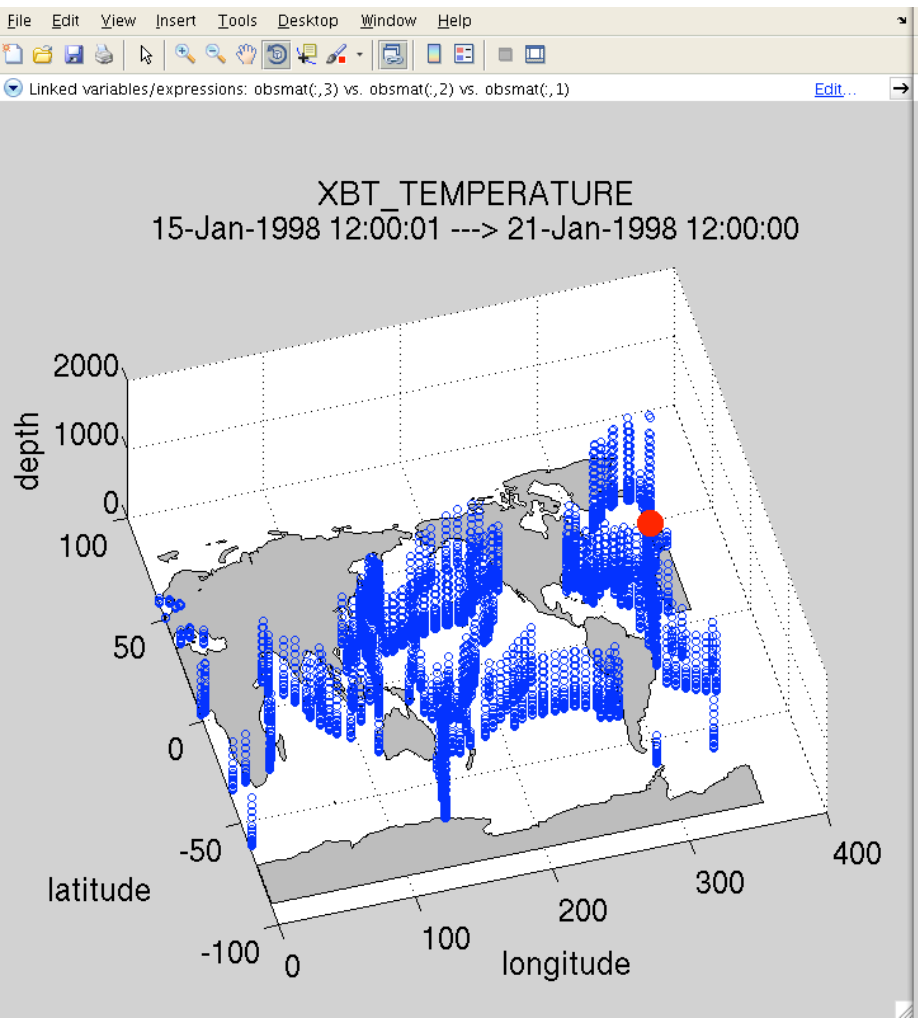
- Each edge shared by two cells - no hanging nodes
- Allows smooth refinement to mitigate many refinement problems

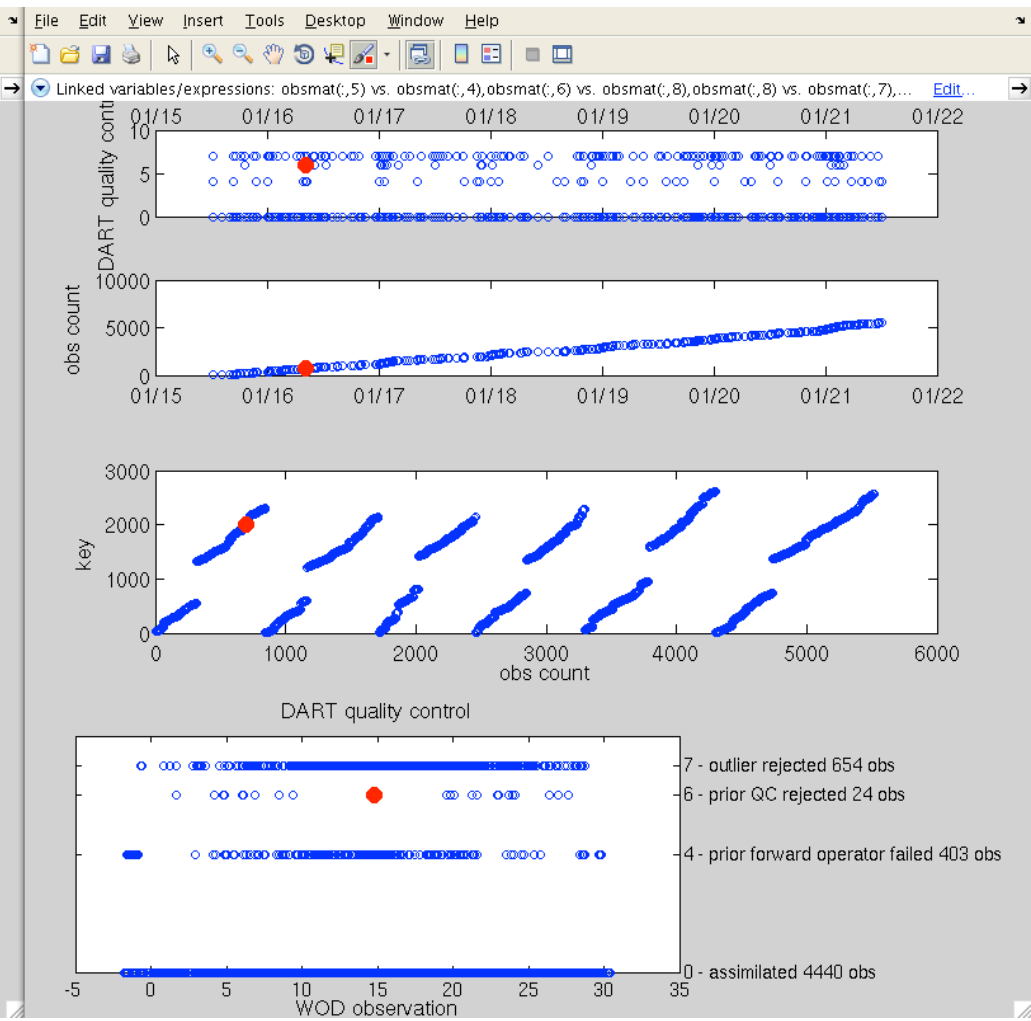


C-grid staggering

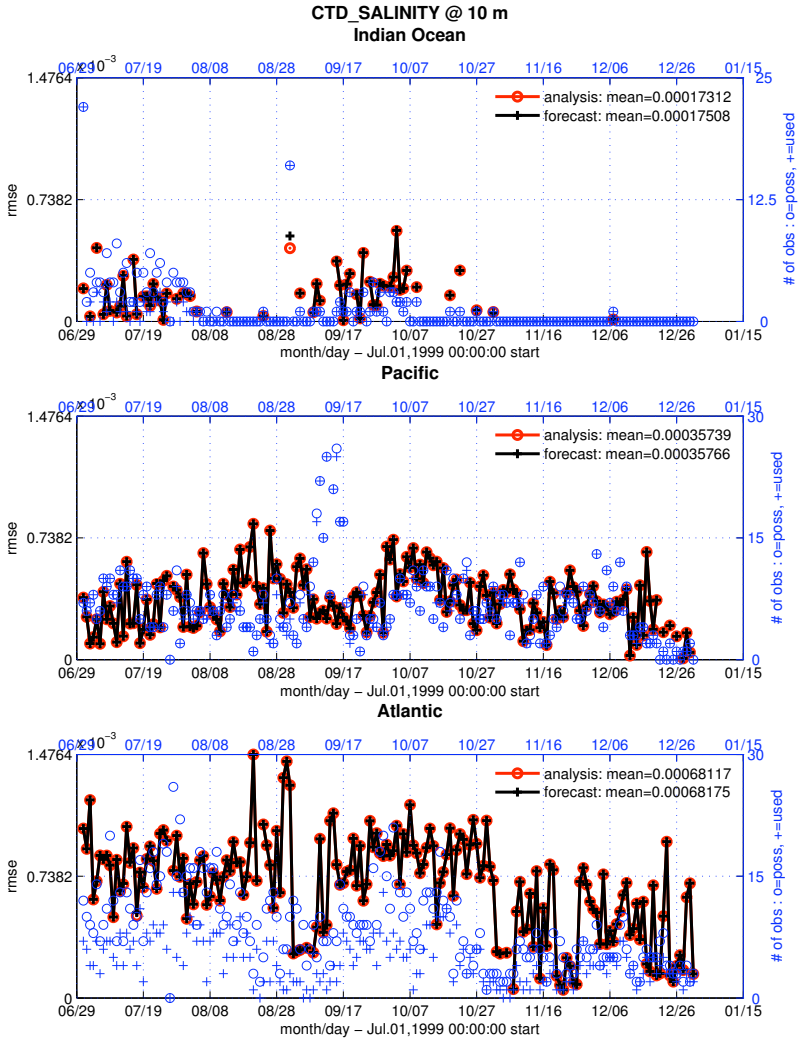
- Provides good accuracy for the fast (gravity-wave) modes
- Avoids the A-grid parasitic mode
- Proper reconstruction of tangential velocity ensures stationary geostrophic modes (Thuburn et al, JCP 2009)

Observation Visualization Tools





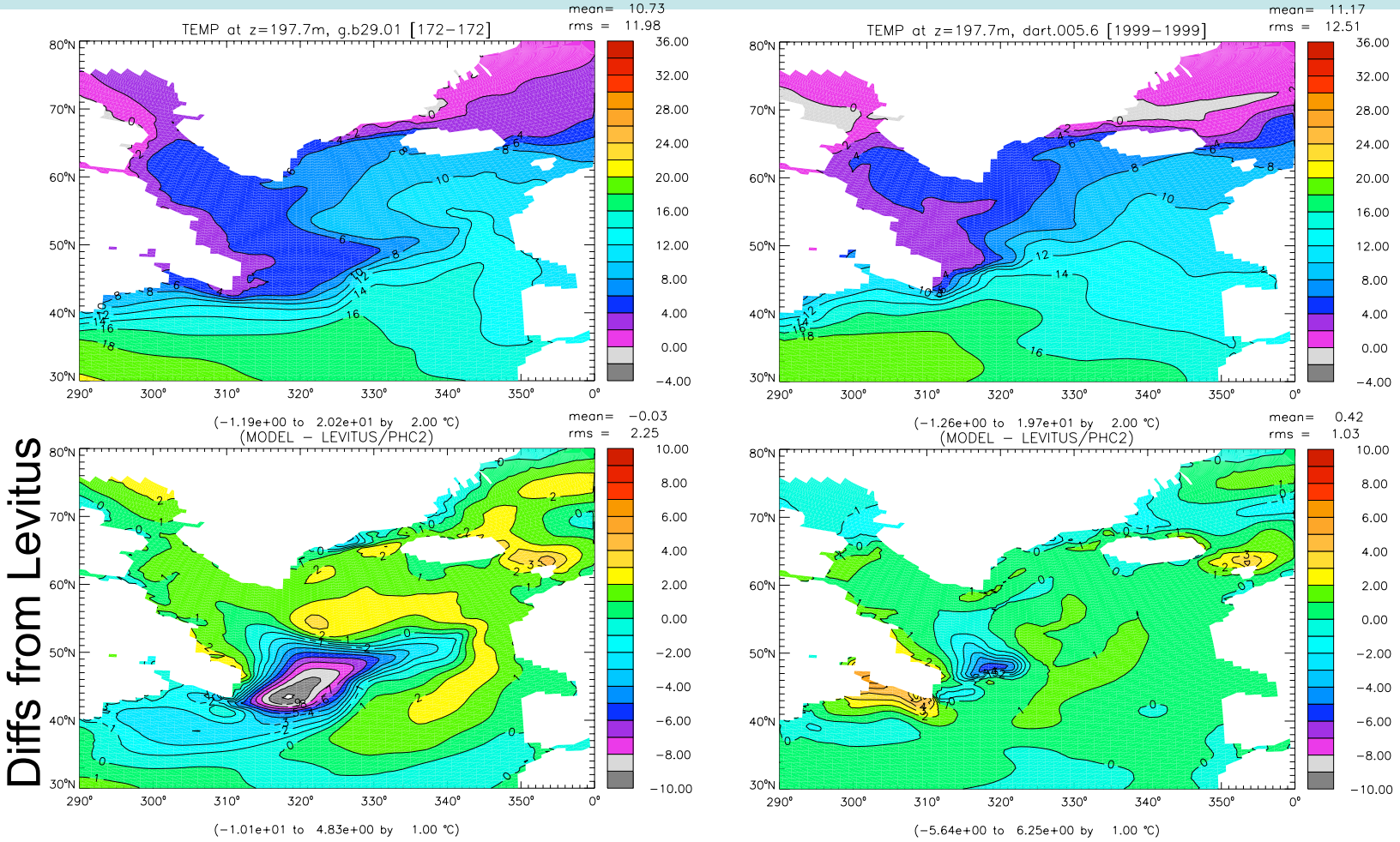
Observation Space Diagnostics (July-Dec. 1999)



10m CTD Salinity

1. Ensemble mean analysis difference from obs
2. Ensemble mean 1-day forecast difference from obs
3. Blue circle is # of obs
4. Blue + is # assimilated
5. Obs. are rejected if they are too far from ensemble mean (3 standard deviations here)

Physical Space Preview: 200m Temperature Means



Our business model

- DART is public domain, NCAR base funded
- The DART group provides consulting, support and implementation help
- Train our customers who are model or observation experts to use DART; keep core group small
- We often host a visit from collaborators as new models are added

DART works with many geophysical models

Global Atmosphere models:

CAM	Community Atmosphere Model; (all 3 dynamical cores)	NCAR
CAM/CHEM	CAM with Chemistry	NCAR
WACCM	Whole Atmosphere Community Climate Model	NCAR
AM2	Atmosphere Model 2	NOAA/GFDL
NOGAPS	Navy Operational Global Atmospheric Prediction System	US Navy
ECHAM	European Centre Hamburg Model	Hamburg
Planet WRF	Global version of WRF	JPL
MPAS	Model for Prediction Across Scales	NCAR/DOE

DART works with many geophysical models

Regional Atmosphere models:

WRF/ARW	Weather Research and Forecast Model	NCAR
WRF/CHEM	WRF with Chemistry	NCAR
NCOMMAS	Collaborative Model for Multiscale Atmospheric Simulation	NOAA/NSSL
COAMPS	Coupled Ocean/Atmosphere Mesoscale Prediction System	US Navy
CMAQ	Community Multi-scale Air Quality	EPA
COSMO	Consortium for Small-Scale Modeling	DWD

DART users work with many observational datasets Atmospheric Observations (1)

U,V,T,Q	NCEP: Radiosonde, AIRCRAFT (commercial), ACARS	BUFR
U,V	NCEP: Cloud Drift Winds from satellite	BUFR
U,V (ocean surface)	QUIKSCAT, including L2B (JPL)	HDF-4
T,Q,refractivity of the atmosphere	COSMIC Global Positioning Satellite radio occultation	NetCDF
T,Q,Tsurface	AIRS from Aqua/A-train satellite	HDF-4, HDF-EOS
U,V,T,Q,Tsurface, pressure,altimeter	MADIS: ACARS, Marine and MESONET surface, METAR, radiosonde, satellite wind	NetCDF
Radar reflectivity, radial velocity	NCEP	Level2 (binary)

DART users work with many observational datasets Atmospheric Observations (2)

U,V	MADIS; Wind Profilers, Atmospheric Motion Vectors (AMVs)	NetCDF, ASCII Text
U,V,T,Q,altimeter	OK mesonet (U. OK)	ASCII Text
Cloud Liquid Water Path, Cloud Top and Base Pressures	GOES satellite, CIMSS	NetCDF
U,V	SSEC (U Wisconsin): Cloud Drift Winds from satellite	ASCII Text
CO (carbon monoxide)	MOPITT	HDF
U,V	GOES CIMSS (U. WI); rapid-scan AMVs (Atmospheric Motion Vectors), satellite cloud winds	CIMSS ASCII

DART users work with many observational datasets Atmospheric Observations (3)

T,Q,Total Precipitable Water	GOES CIMSS hyperspectral AIRS IR	CIMSS ASCII
Total Precipitable Water	AMSR, MODIS Microwave	ASCII Text
U,V	Operational typhoon bogus winds, Taiwan Central Weather Bureau	ASCII Text
U,V (at wind turbine hub height)	Seimens(?)	?
Electron density	COSMIC/FORMOSAT-3	LDM (UCAR/Unidata)
U,V,T	GTS	little-r
Chemical concentrations	IASI on EUMETSAT Polar System MetOp satellite	converted to ASCII intermediate format
Aerosol optical depth (AOD)	TERA and AQUA	HDF

DART works with many geophysical models

Upper Atmosphere/Space Weather models:

ROSE		NCAR
TIEGCM	Thermosphere Ionosphere Electrodynamic GCM	NCAR/HAO
GITM	Global Ionosphere Thermosphere Model	Michigan
Solar Dynamo	Dynamo/sunspot model	NCAR/HAO

DART users work with many observational datasets Solar, Space Weather, Extraterrestrial Observations:

Radiances, Occultation on Mars	TES, limb sounder on Mars	?
Density, ion concentrations	CHAMP	NetCDF
Thermospheric Mass Densities	CHAMP, GRACE	NetCDF
Electron densities	COSMIC	NetCDF
Total Electron Density	Garner GPS Archive	RINEX
Orbital element information	NORAD	ASCII
Solar Magnetic Fields	Wilcox, Mt Wilson, National Solar Observatories	?
Rotational, Meridional Circulation	Mt Wilson, SoHO, SDO, HMI	?

DART works with many geophysical models

Land Surface models:

CLM	Community Land Model	NCAR
NOAH	Relatively simple land model	Community

DART users work with many observational datasets
Land Observations:

Snow cover	MODIS	HDF
Heat Flux, Net Carbon	Ameriflux tower network	ASCII Text
Soil Moisture	COSMOS (neutron counter)	ASCII Text

DART works with many geophysical models

Ocean models:

POP	Parallel Ocean Program	DOE/NCAR
MIT OGCM	Ocean General Circulation Model	MIT
ROMS	Regional Ocean Modeling System (under development)	Rutgers
MPAS	Model for Prediction Across Scales (Under development)	DOE/LANL

DART users work with many observational datasets Ocean Observations:

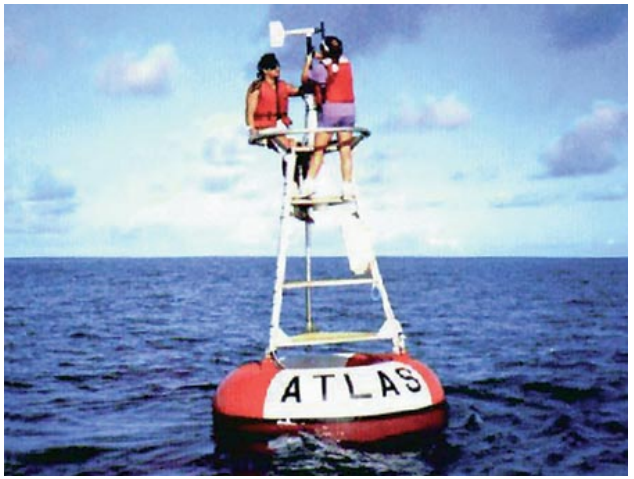
T, Salinity	World Ocean Database: Argo floats, CTD(ships), XBT, moored thermistors, drifting buoys(GT-SPP)	packed ASCII
Surface U, V currents	CODAR	ASCII Text

Ocean Model State

- "State" is the minimal set of values required to restart the model
  - Ignoring derived values
  - Modified by the assimilation
- Example from the "POP" ocean model:
  - Potential Temperature
  - Salinity
  - U, V Currents
  - Sea Surface Height

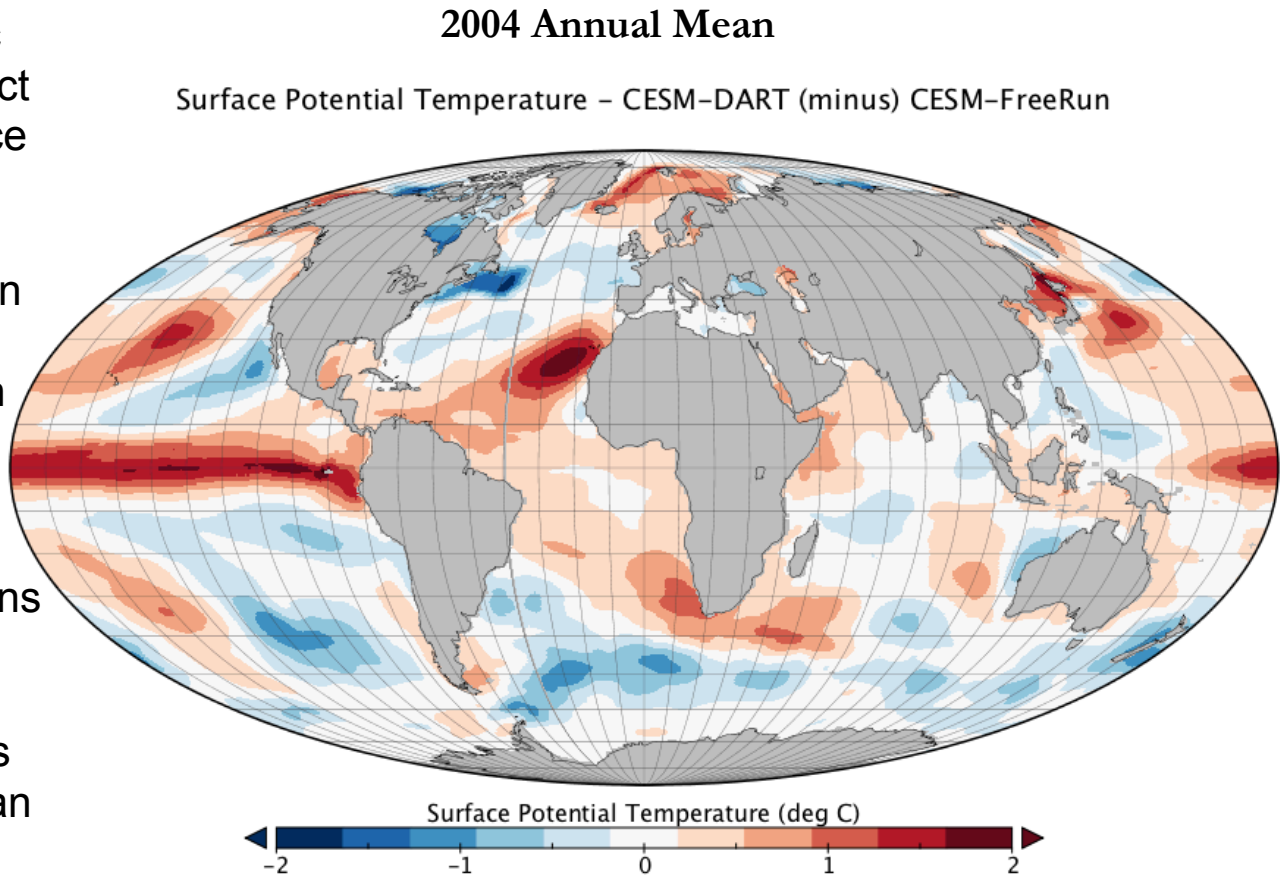
Temperature,Salinity from World Ocean Database 2005

FLOAT_SALINITY	68200
FLOAT_TEMPERATURE	395032
DRIFTER_TEMPERATURE	33963
MOORING_SALINITY	27476
MOORING_TEMPERATURE	623967
BOTTLE_SALINITY	79855
BOTTLE_TEMPERATURE	81488
CTD_SALINITY	328812
CTD_TEMPERATURE	368715
STD_SALINITY	674
STD_TEMPERATURE	677
XCTD_SALINITY	3328
XCTD_TEMPERATURE	5790
MBT_TEMPERATURE	58206
XBT_TEMPERATURE	1093330
APB_TEMPERATURE	580111



Recent Results from CESM-DART simulation

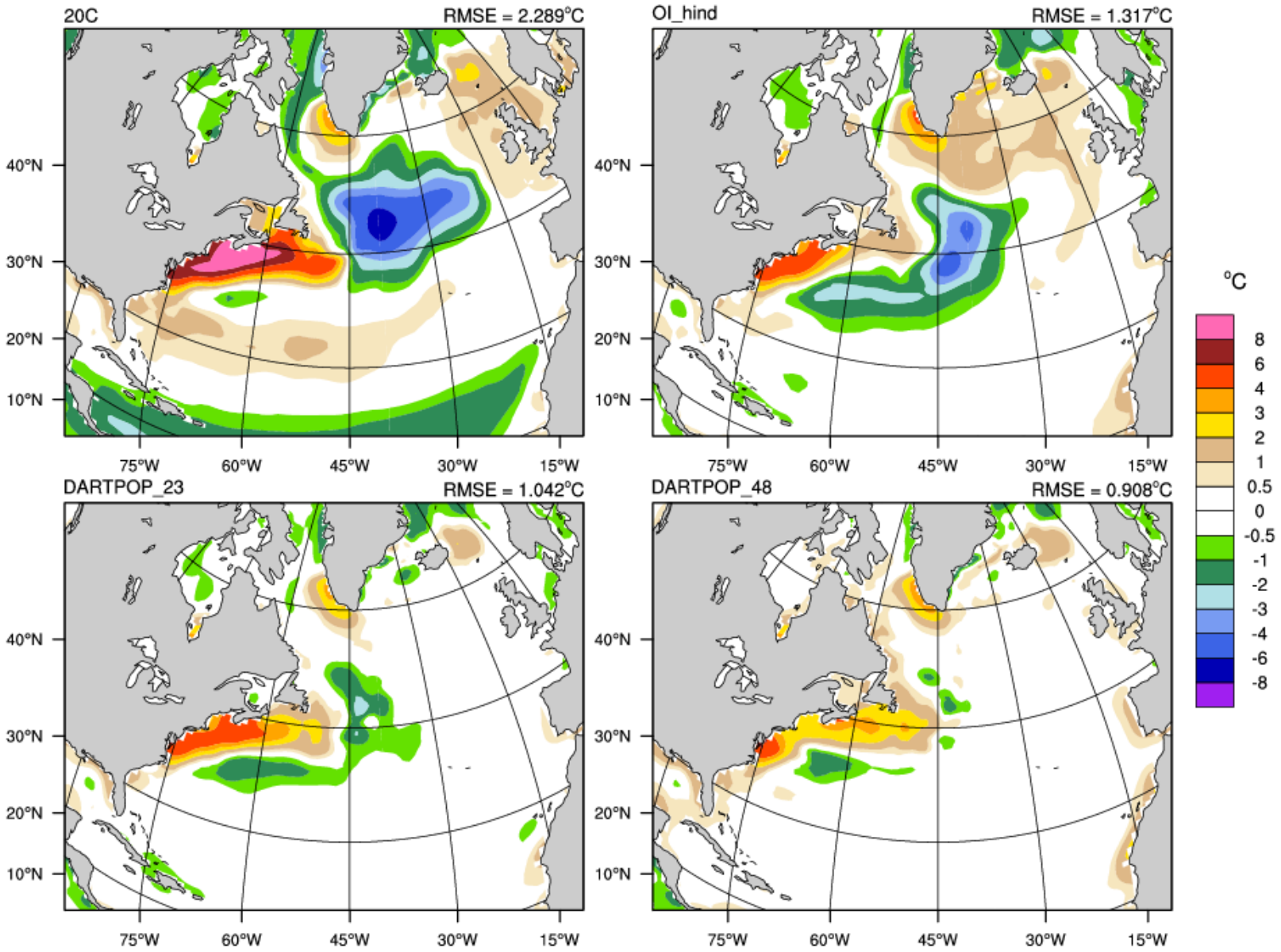
- Assimilation of atmospheric and oceanic observations make distinct adjustments to the surface potential temperature
- Pronounced differences in the Tropics, Western boundary currents, South Atlantic/South Indian
- Comparison to unassimilated observations show the CESM-DART estimate of surface potential temperature has lower RMSE and bias than a free run of CESM



Physical Space: 1998/1999 SST Anomaly from HadOI-SST

Coupled Free Run

23 POP, 1 DATM



POP forced by observed atmosphere

48 POP, 48 DATM

DART Evolution Challenges

- DART runs well on O(10 – 1000) processors
- New architectures O(100,000) processors
- Highly scalable systems require less global communication, more asynchronicity
  - Less memory per node, more nodes, lower power
  - Harder to program Geophysical applications

Addressing Shrinking Memory Sizes

- Redesigning forward operator algorithms to avoid the need for entire state of one ensemble member in single task memory
- Requires additional communication for some types of forward operators
- Keeping spatial locality lowers communication overhead but presents load balancing issues

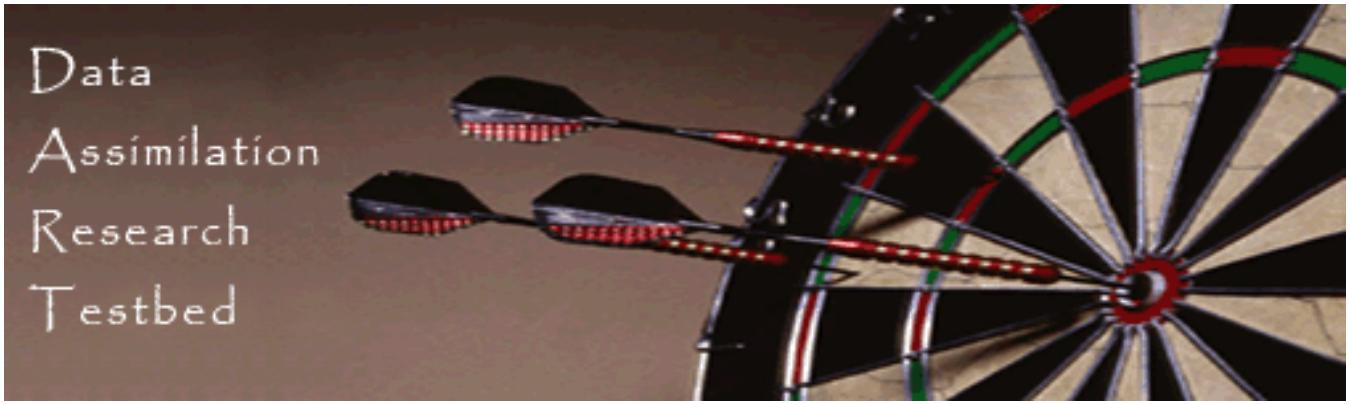
Avoiding Global Communication

- Current implementation transposes data for load balancing during state adjustment phase
- Global operations prohibitively expensive on O(100,000) processor counts
- Avoiding transposes avoids global operation but again raises more load balancing issues

DART Evolution for MPP Systems

- Allow single ensemble state to span multiple tasks
  - Decompose across a small number of nodes
  - Data movement confined to subsets of nodes
- Support distributed forward operator computations
  - Spatially local decomposition minimizes communication
  - One-sided MPI-2 communication avoids barriers
- Avoid global communication at state adjustment phase
  - Smarter decomposition for load balancing
  - Parallel adjustments of disjoint observation sets

Thank You



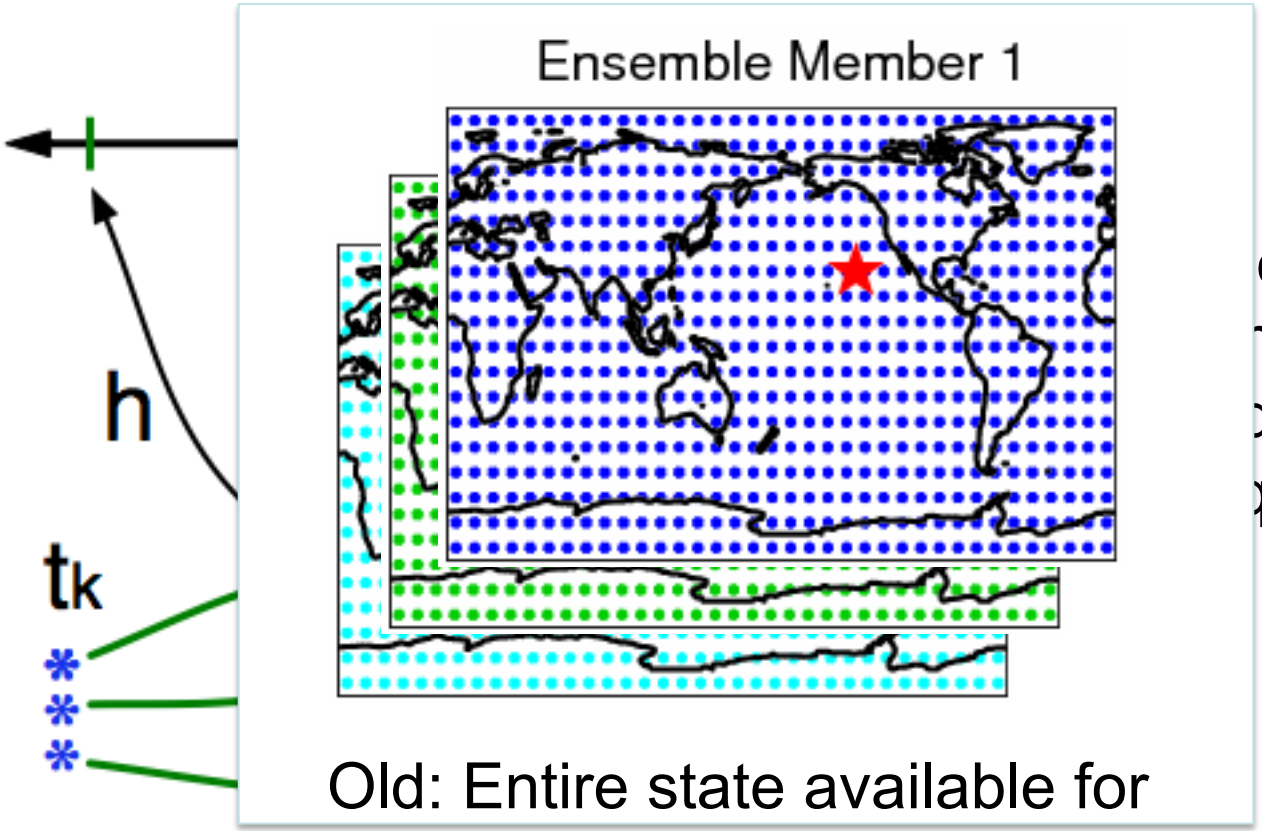
www.image.ucar.edu/DAReS/DART

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Arellano, A., 2009: *The Data Assimilation Research Testbed: A community facility.* BAMS, **90**, 1283—1296, doi: 10.1175/2009BAMS2618.1


Data
Assimilation
Research
Testbed

Ensemble Filter For Large Geophysical Models

2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.

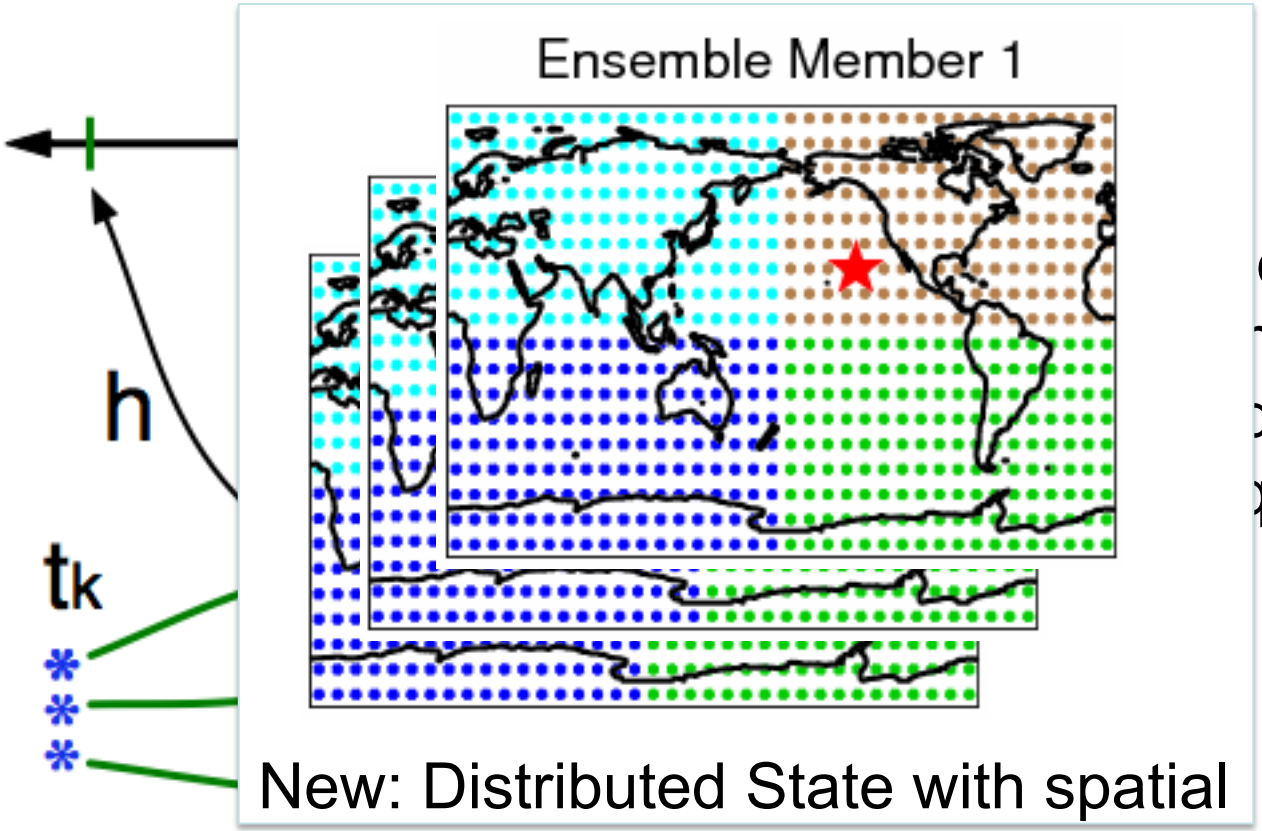


ervations
nents with
d errors can
quentially.

Old: Entire state available for forward ops

Ensemble Filter For Large Geophysical Models

2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.

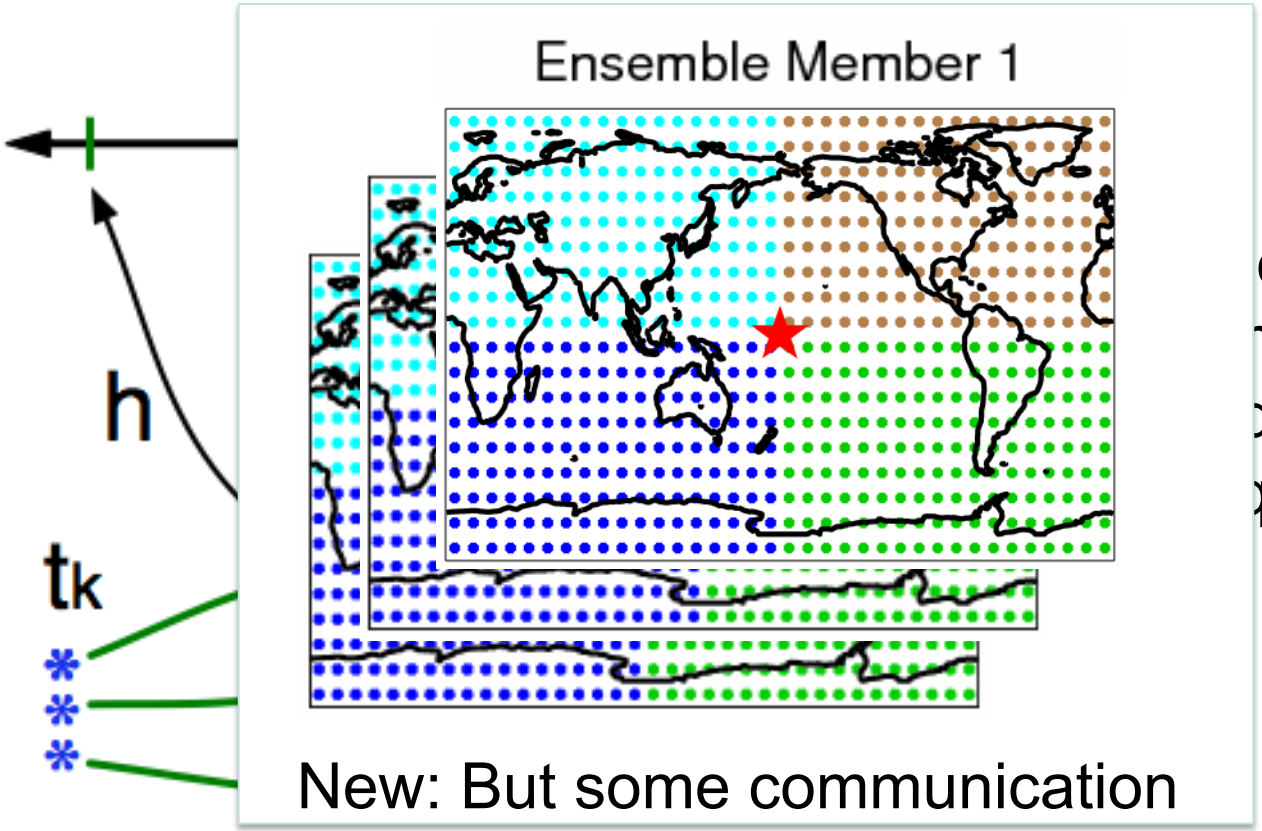


ervations
nents with
d errors can
quentially.

New: Distributed State with spatial locality

Ensemble Filter For Large Geophysical Models

2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.

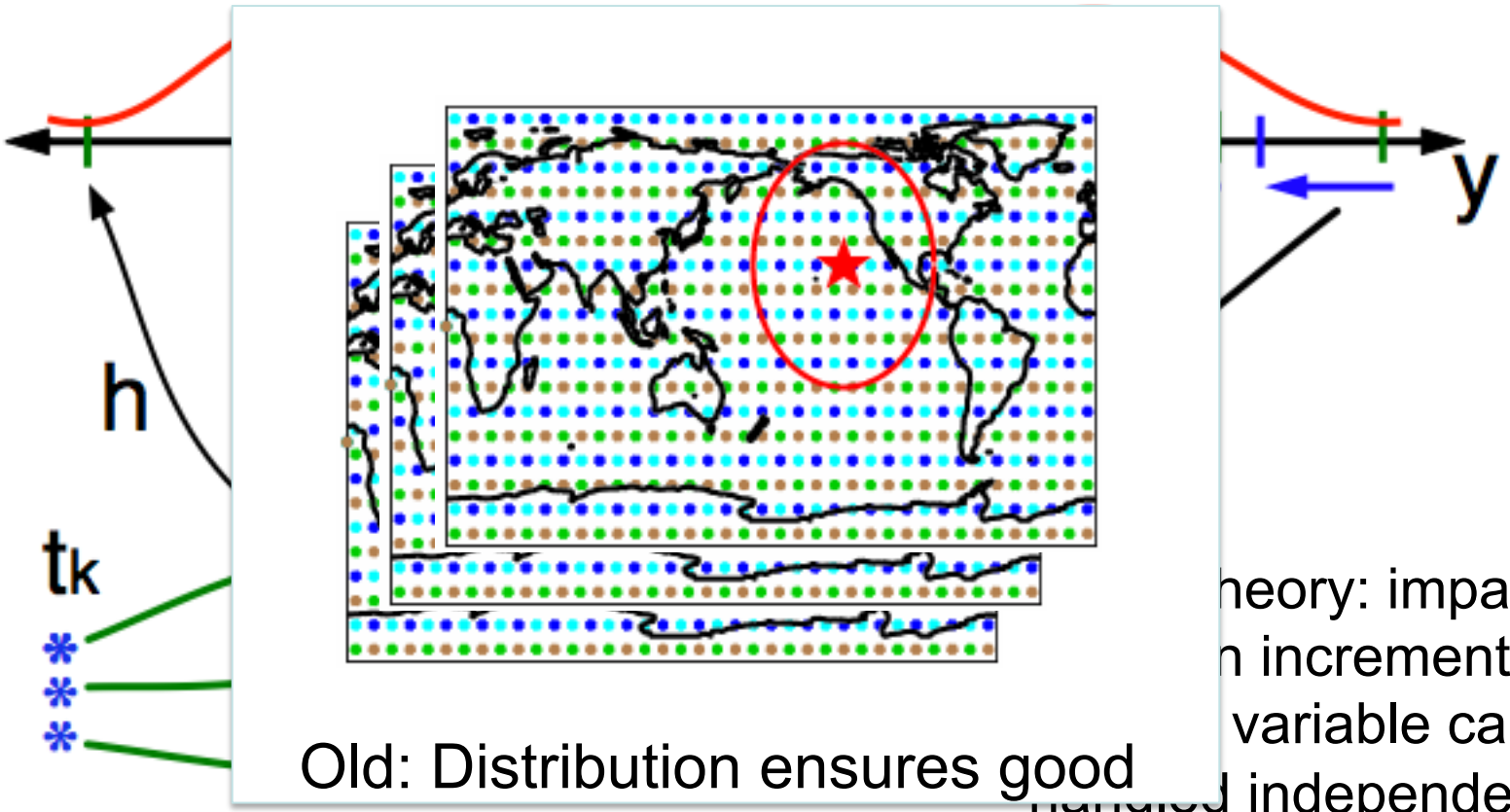


ervations
nents with
d errors can
quentially.

New: But some communication may be needed

Ensemble Filter For Large Geophysical Models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.

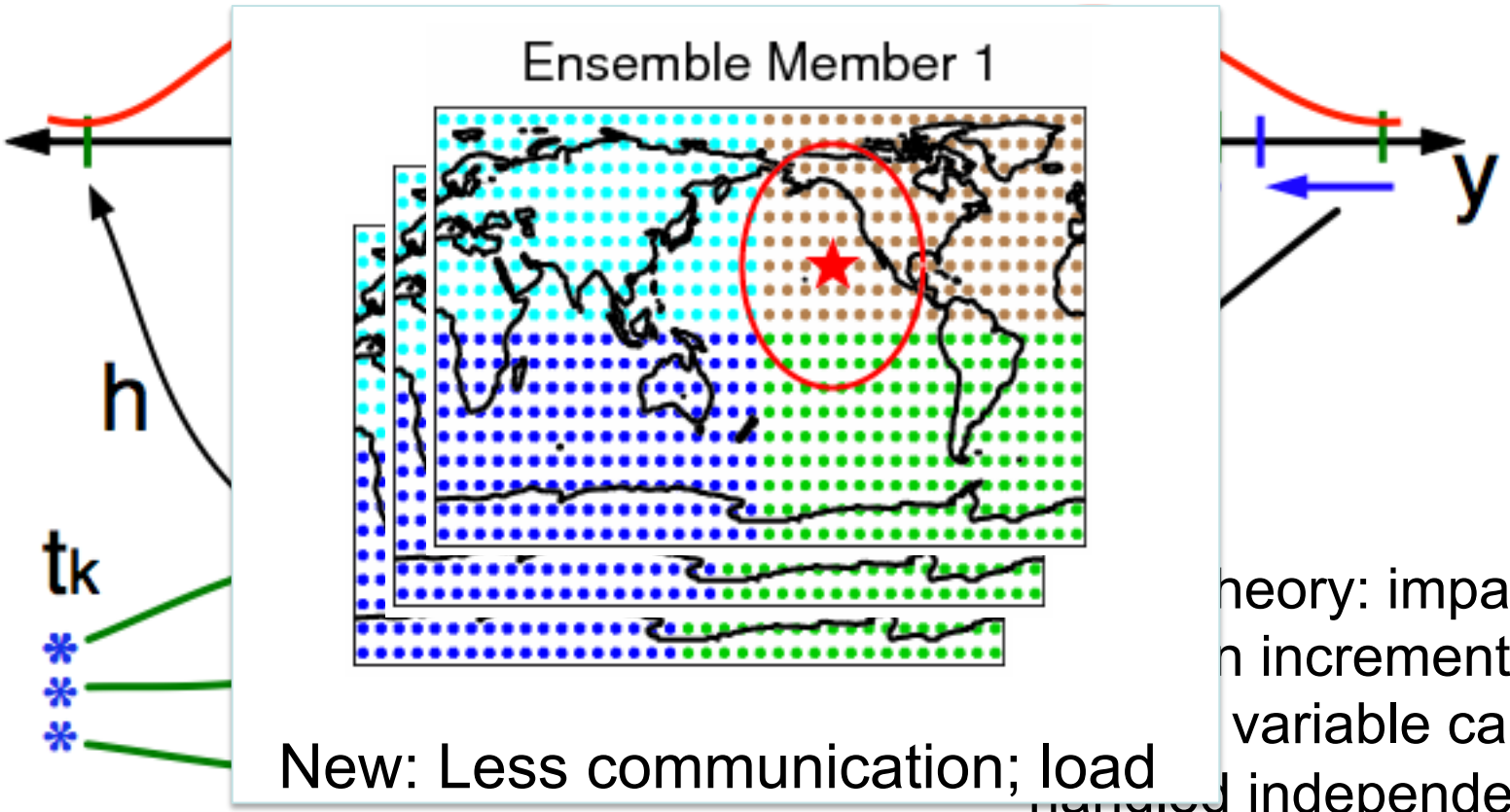


Old: Distribution ensures good load balancing

heory: impact of
n increments on
variable can be
d independently!

Ensemble Filter For Large Geophysical Models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.

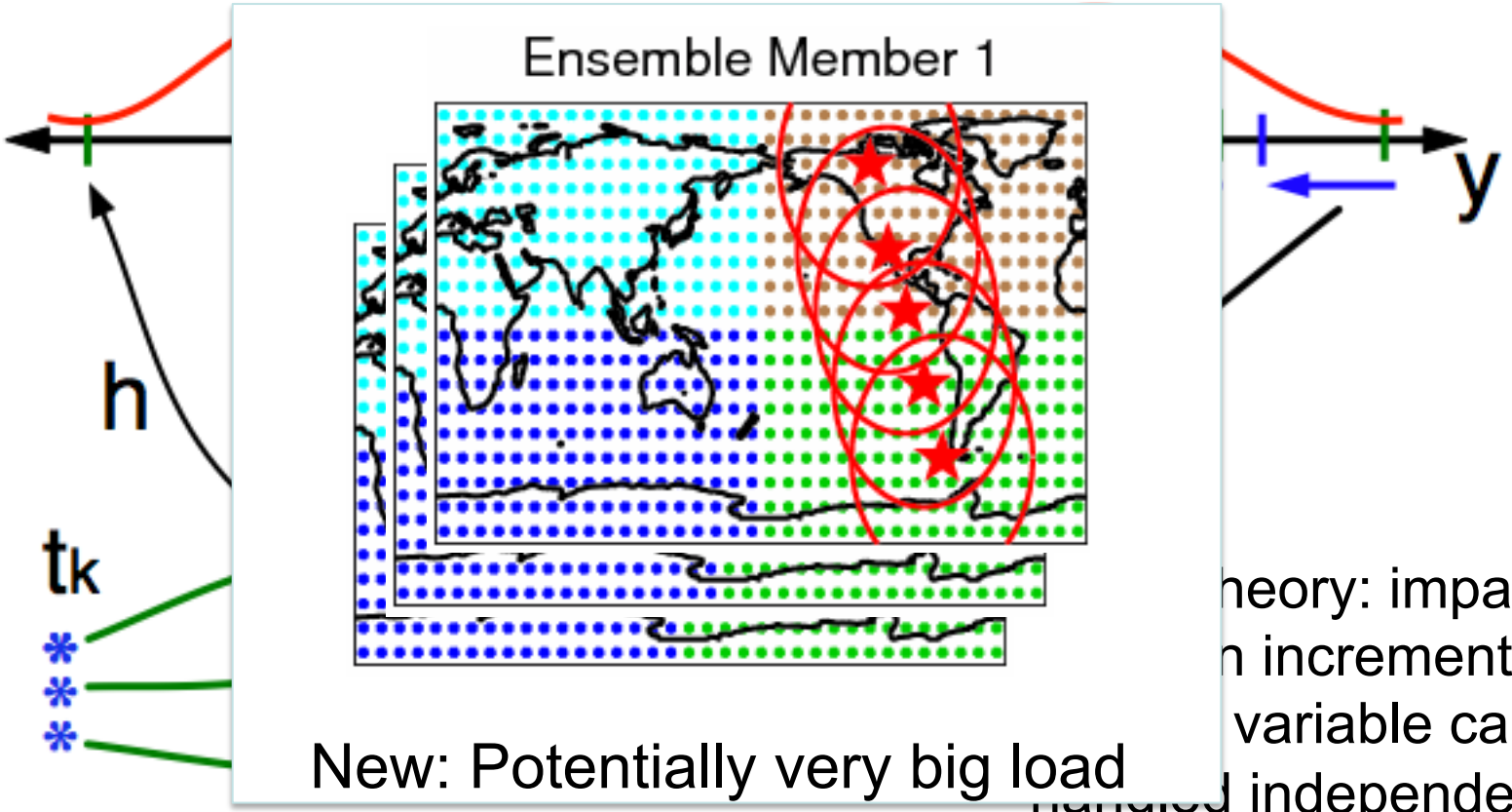


heory: impact of
n increments on
variable can be
independently!

New: Less communication; load balance issues

Ensemble Filter For Large Geophysical Models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.

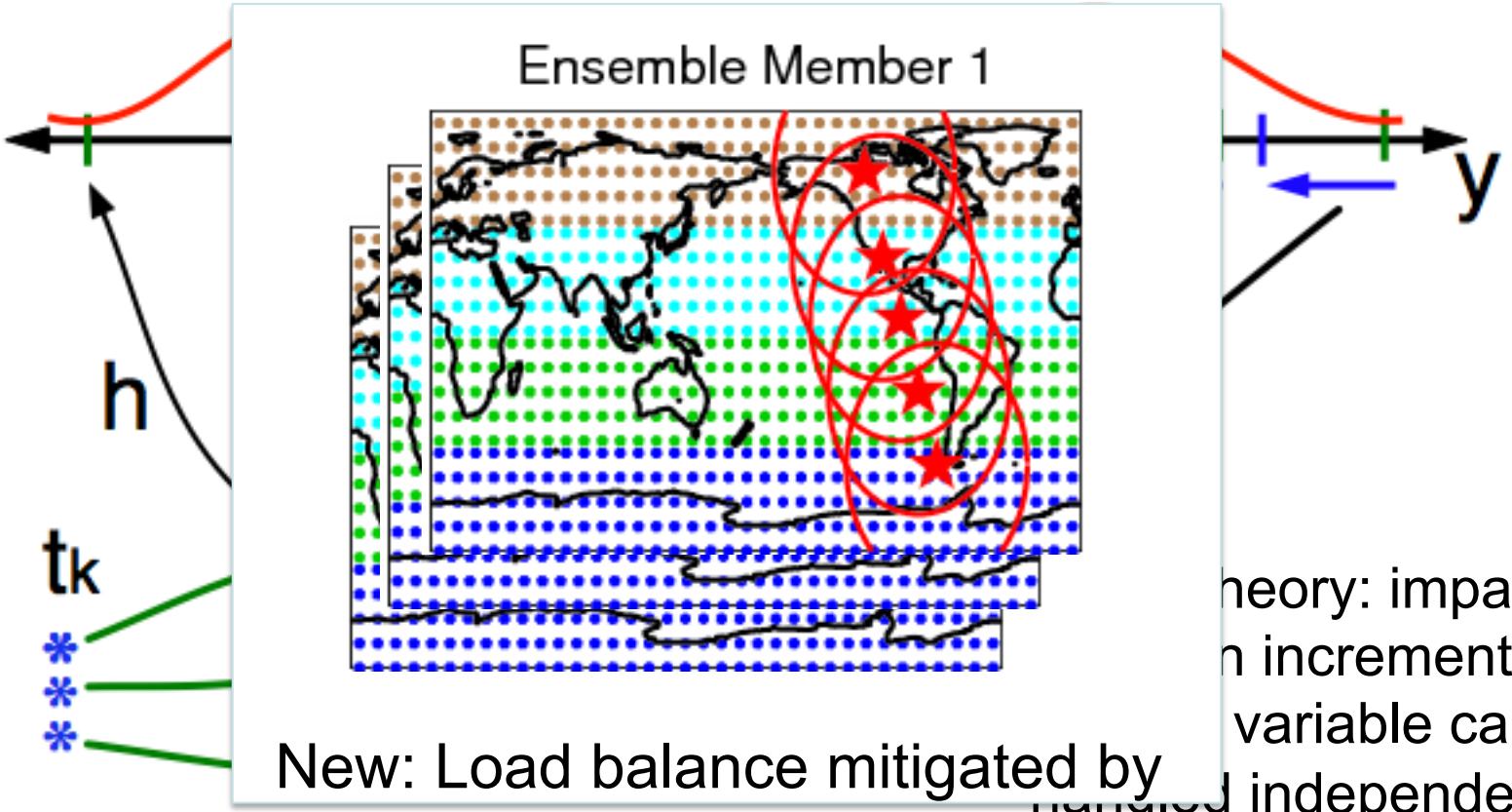


heory: impact of
n increments on
variable can be
d independently!

New: Potentially very big load balance issues

Ensemble Filter For Large Geophysical Models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.

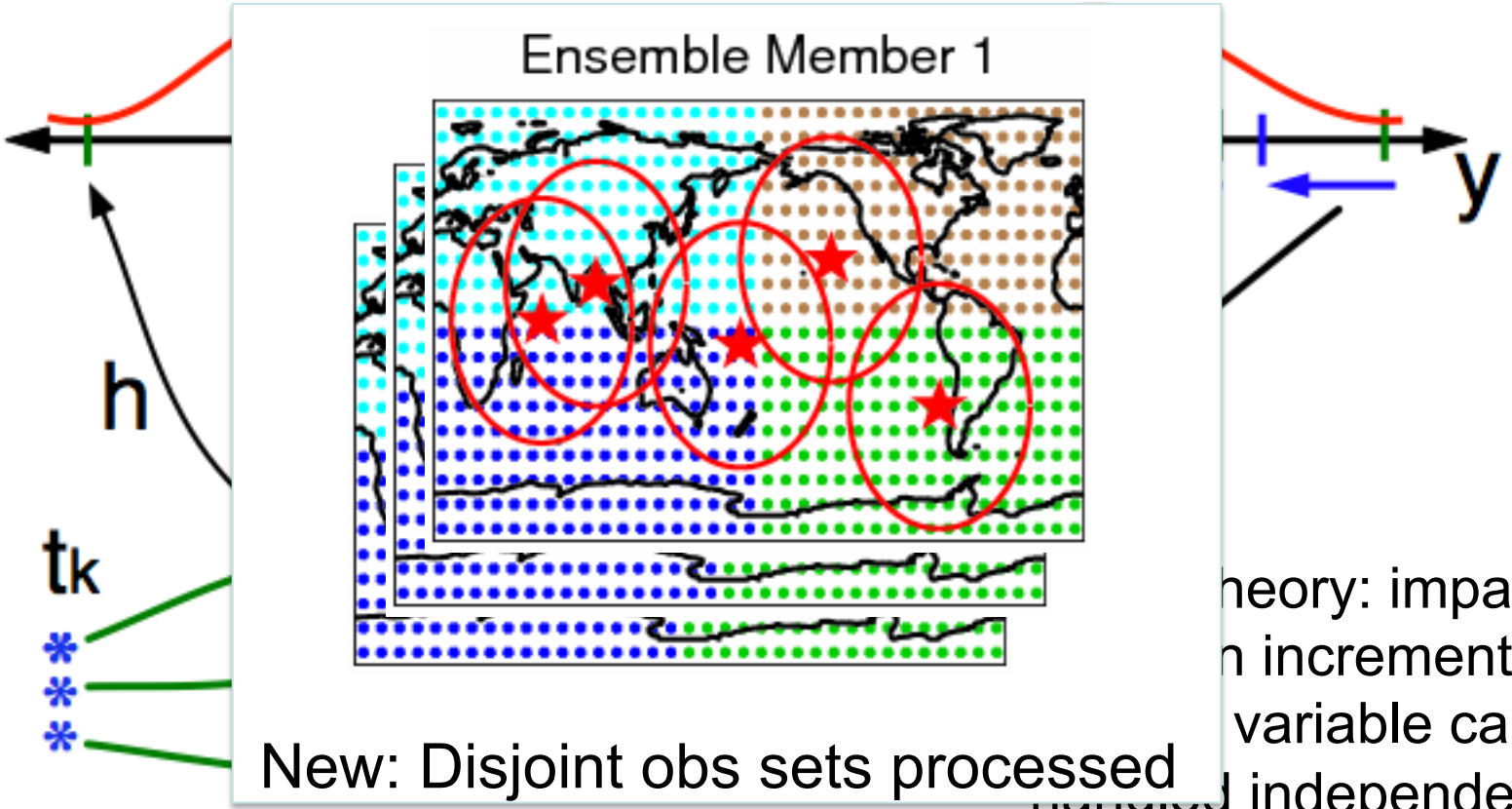


heory: impact of
n increments on
variable can be
d independently!

New: Load balance mitigated by layout choice

Ensemble Filter For Large Geophysical Models

5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



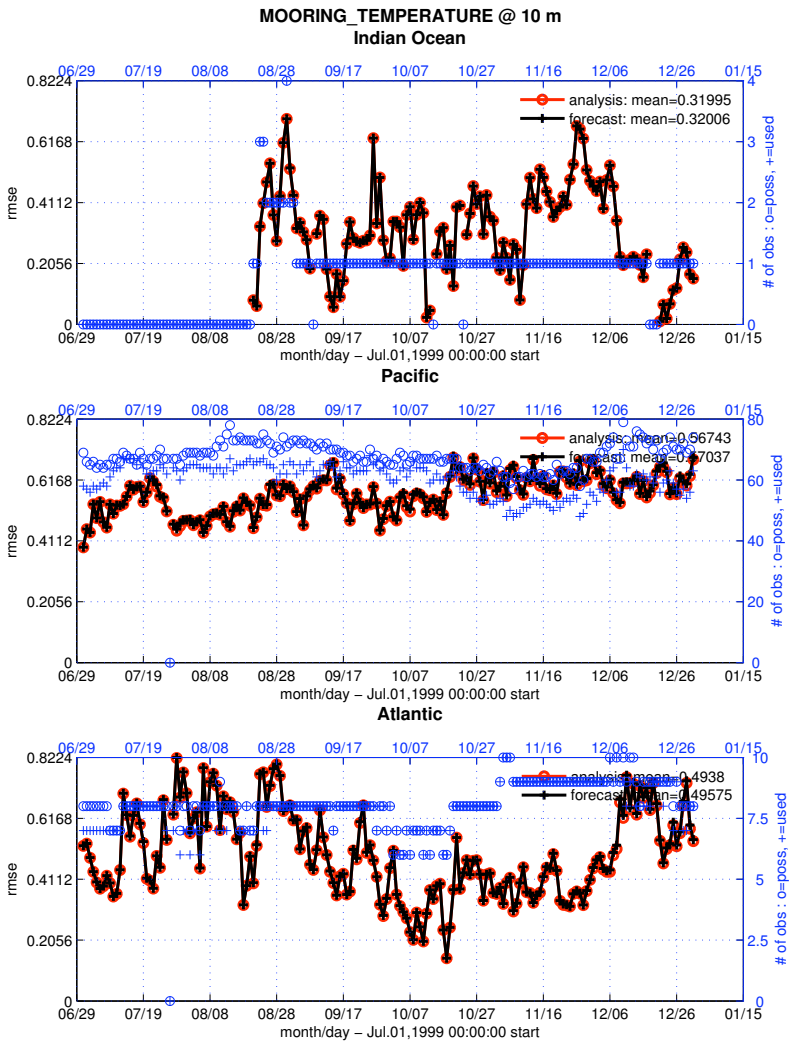
DART Evolution (cont)

Maintain reasonable interfaces that enable user-extensible sections of the code

- Support for modification by domain scientists
- Clear and understandable process for adding new models and new observation operators
- Encapsulate MPI code at a level where user does not have to understand the details

Transformational hardware architecture changes may require transformational algorithmic choices

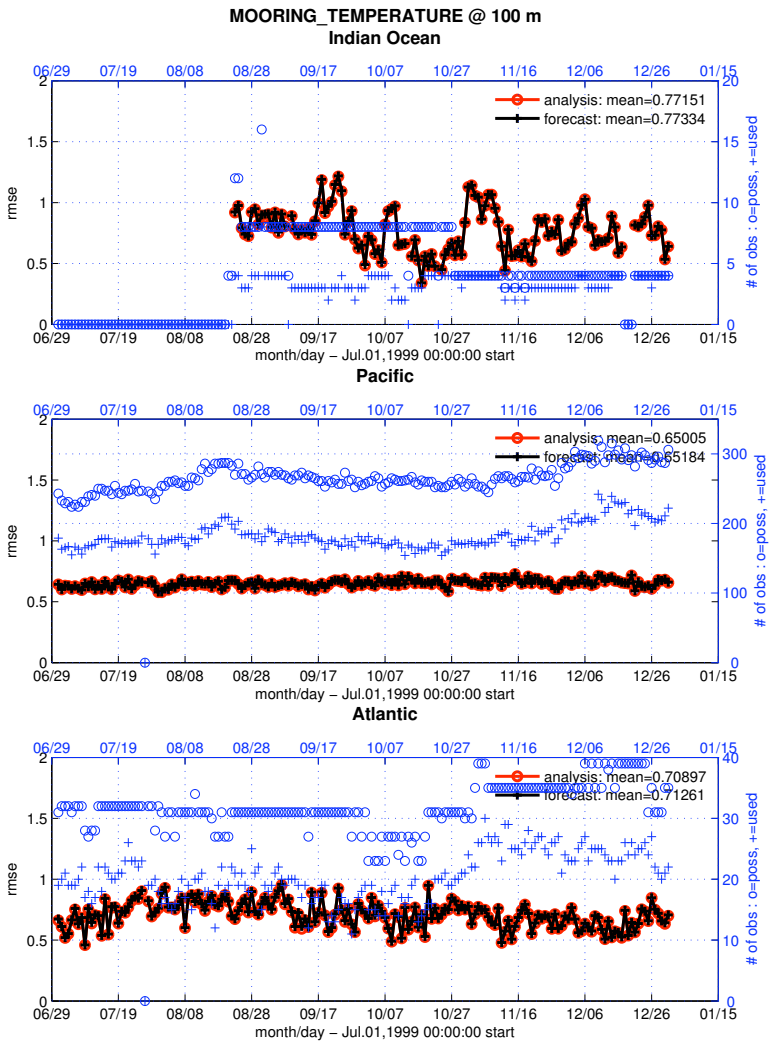
Observation Space Diagnostics (July-Dec. 1999)



10m Mooring Temperature

1. Ensemble mean analysis difference from obs
2. Ensemble mean 1-day forecast difference from obs
3. Blue circle is # of obs
4. Blue + is # assimilated
5. Obs. are rejected if they are too far from ensemble mean (3 standard deviations here)

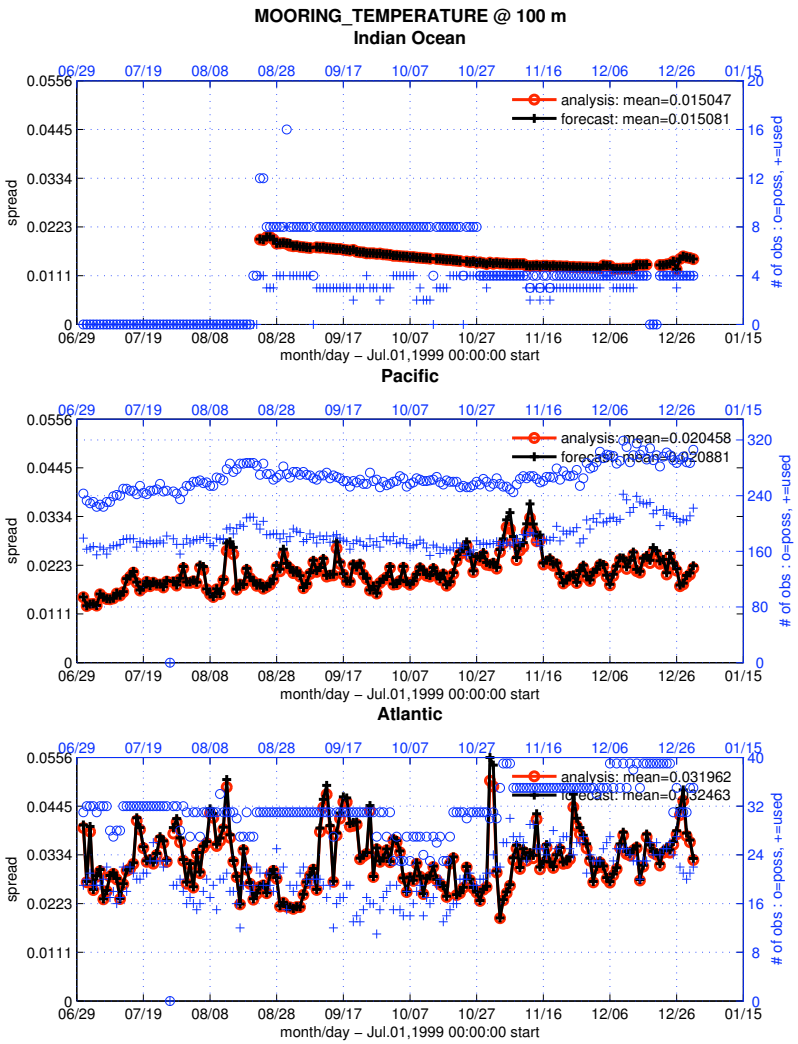
Observation Space Diagnostics (July-Dec. 1999)



100m Mooring Temperature

1. Blue circle is # of obs.
2. Blue + is # assimilated.
3. Obs. are rejected if they are too far from ensemble mean (3 standard deviations here).
4. About 1/3 of obs. rejected.
5. Model bias in thermocline?

Observation Space Diagnostics: Ensemble Spread



100m Mooring Temperature

1. Spread is way too small
2. Model bias makes this even worse
3. Using single atmospheric forcing is part of the problem
4. Automatic spread correction tools in DART weren't working with POP at this time