

## Introduction to ASP 2003 Colloquium Exercises

Jeffrey Anderson  
NOAA/GFDL and NCAR/DAI

Colloquium sponsored by NCAR's ASP, MMM, CGD and DAI

DAI (Data Assimilation Initiative) formed 2002 to focus NCAR research and development activities

“DAI aims to create and lead a research community for data assimilation where individuals benefit from sharing ideas, methodologies, and software tools as well as access to a data assimilation testbed.”

## Goals: DAI will provide:

1. A data assimilation community within NCAR to produce leading-edge research and to provide focus to disparate efforts;

## Goals: DAI will provide:

1. A data assimilation community within NCAR to produce leading-edge research and to provide focus to disparate efforts;
2. A software environment for supporting data assimilation research and evaluation; the Data Assimilation Research Testbed (DART);

## Goals: DAI will provide:

1. A data assimilation community within NCAR to produce leading-edge research and to provide focus to disparate efforts;
2. A software environment for supporting data assimilation research and evaluation; the Data Assimilation Research Testbed (DART);
3. A mechanism for data assimilation research collaboration with strategically selected partners from universities and government research labs;

## Goals: DAI will provide:

1. A data assimilation community within NCAR to produce leading-edge research and to provide focus to disparate efforts;
2. A software environment for supporting data assimilation research and evaluation; the Data Assimilation Research Testbed (DART);
3. A mechanism for data assimilation research collaboration with strategically selected partners from universities and government research labs;
4. Software tools for use in undergraduate and graduate education;

## Goals: DAI will provide:

1. A data assimilation community within NCAR to produce leading-edge research and to provide focus to disparate efforts;
2. A software environment for supporting data assimilation research and evaluation; the Data Assimilation Research Testbed (DART);
3. A mechanism for data assimilation research collaboration with strategically selected partners from universities and government research labs;
4. Software tools for use in undergraduate and graduate education;
5. Basic research and implementation support for 'operational' partners, both within NCAR and outside.

## Goals for this workshop:

1. Attract high quality students to data assimilation research

## Goals for this workshop:

1. Attract high quality students to data assimilation research
2. Help prepare these students to tackle difficult 'big' problems

## Goals for this workshop:

1. Attract high quality students to data assimilation research
2. Help prepare these students to tackle difficult 'big' problems
3. Test early versions of Data Assimilation Research Testbed

## Goals for this workshop:

1. Attract high quality students to data assimilation research
2. Help prepare these students to tackle difficult 'big' problems
3. Test early versions of Data Assimilation Research Testbed
4. Begin building a user community for DART

## Goals for this workshop:

1. Attract high quality students to data assimilation research
2. Help prepare these students to tackle difficult ‘big’ problems
3. Test early versions of Data Assimilation Research Testbed
4. Begin building a user community for DART
5. Learn more about students’ research activities in DA

## Goals for this workshop:

1. Attract high quality students to data assimilation research
2. Help prepare these students to tackle difficult ‘big’ problems
3. Test early versions of Data Assimilation Research Testbed
4. Begin building a user community for DART
5. Learn more about students’ research activities in DA
6. Finish with students and lecturers excited about Data Assimilation

## Reaching these goals:

1. Most knowledgeable, best presenters from throughout operational DA community
2. Talks ranging from basics to cutting edge research topics
3. Interactive exercise sets designed to help students think about DA concepts
4. Environment that encourages students to participate in discussion

## Lecture Schedule

Monday	Tuesday	Wednesday	Thursday	Friday
Global and Regional Atmospheric Models	Introduction to: Statistics, Data Assim., Observations	DA Methods: OI and 3D-Var 4D-Var  Computational Issues	DA Methods: Kalman Filter Ensemble Filter  Quality Control of Observations	Covariance Models  A Posteriori Diagnostics  Model Error
Operational Systems:  NCEP  ECMWF  Canadian	Satellite Observations  Parameter Tuning (Bayesian)  Hybrid Assim. Methods	Smaller Scales  Regional DA  Convective Scale DA	GPS Observations  Chemical Assimilation  Filters and Variational	Student Presentations  What's fun to do next???

## Exercise Schedule

Monday	Tuesday	Wednesday	Thursday	Friday
WRF / MM5 3D-Var Tutorial Web-based, introductory	Introduction to Data Assimilation		4D-Var Data Assimilation in Simple Model	Introduction to DART: Basic Filters in Low- order models
More DART:  Exploring observation characteristics and model bal- ances	More DART:  Bigger models, ensemble size, etc.  Student special interest projects		Complete stu- dent special interest projects	Additional time if needed

Some exercises forced to occur before corresponding lectures  
Unclear how far we'll get with DART exercises, code can travel

## Exercise Philosophy: Simplifying the DA Problem

1. Split into component parts (discussion and software modularity)
  - A. Prediction models
  - B. Observations
  - C. Assimilation algorithms

Advantage: Avoid overwhelming complexity

Risk: Over-simplifying, neglecting complex interactions

## Exercise Philosophy: Simplifying the DA Problem

### 2. Simplify the individual components

#### A. Use low-order (perfect) models

Easier to diagnose and understand

Cheaper to run and store data

Isolate particular aspects of dynamical systems

Avoid unknown model errors

#### B. Use clean, small observation sets; synthetic observations

Avoid complexity of unknown instrument errors

Avoid unknown 'forward operators'

Don't need massive, convoluted data sets

#### C. Clean assimilation algorithms

Don't need contingencies to deal with problems above

## Exercise Philosophy: Simplifying the DA Problem

3. Put the simple pieces back together again

Gradually remove simplifications and observe behavior

Classical 'hierarchical' approach used in modeling studies

Warning: This methodology has not been overwhelmingly successful for DA R&D

Can one generalize from simple cases to vast operational cases?

## Introduction to Exercise Sets:

1. WRF / MM5 3D-Var tutorial (Dale Barker, NCAR/MMM)
2. Interactive introduction to Data Assimilation (Chris Snyder, NCAR/MMM & DAI)
3. 4D-Var in low-order models (Jenny Sun, NCAR/MMM)
4. Data Assimilation Research Testbed (Jeff Anderson, NCAR/MMM)

# The Data Assimilation Research Testbed (DART)

## What is DART?

1. Allows combinations of assimilation algorithms, models, and observation sets
2. Diagnostic tools
3. Supports Data Assimilation R&D for NCAR and external partners

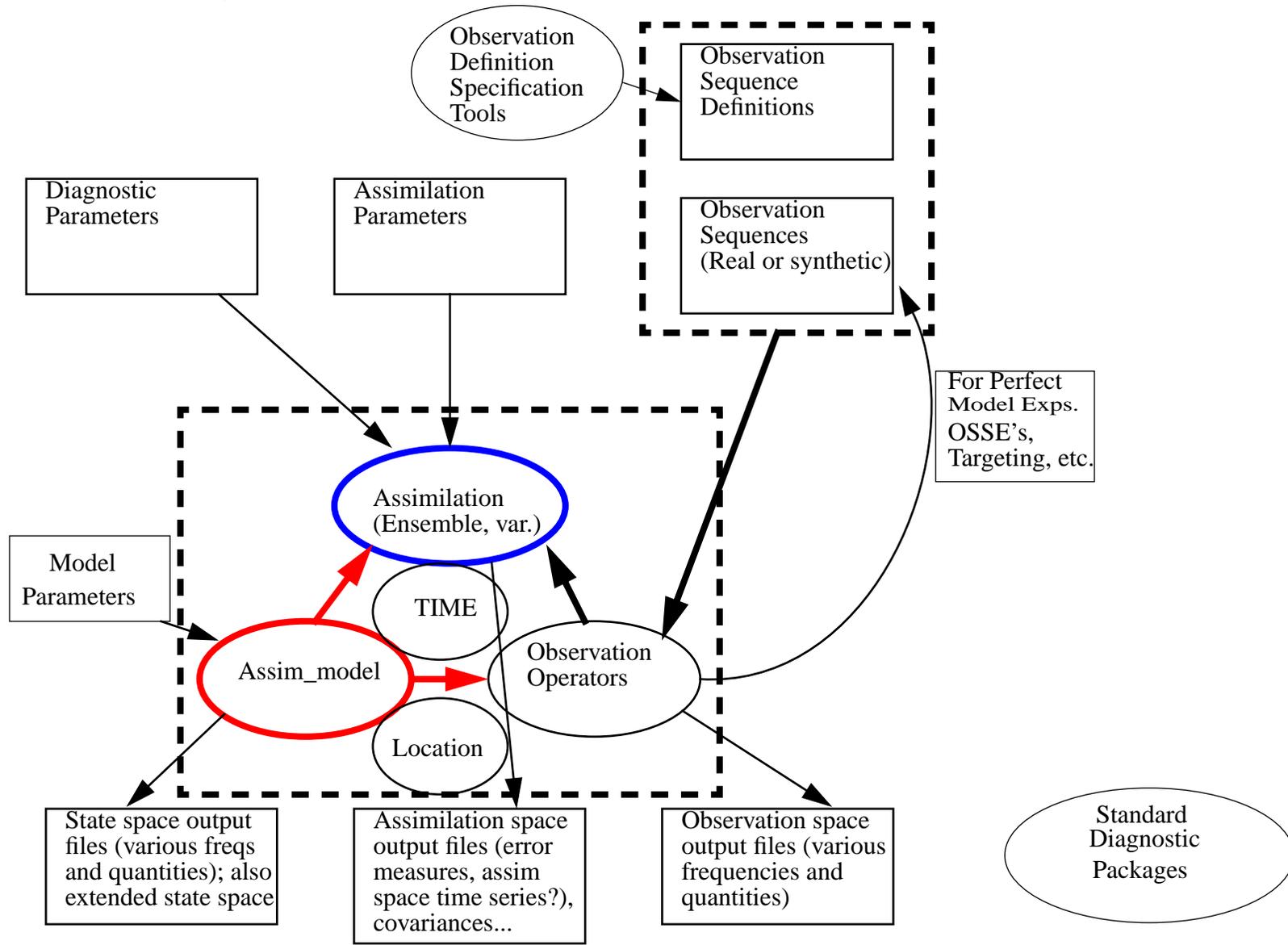
## Status of DART

1. Basic framework implemented
2. Currently using GFDL FMS infrastructure
3. Switch to ESMF (NASA) infrastructure when available
4. Primarily implementing ensemble (Kalman) filters
5. Variational for low-order models only
6. Plans MAY include a variational (4D-Var) capability

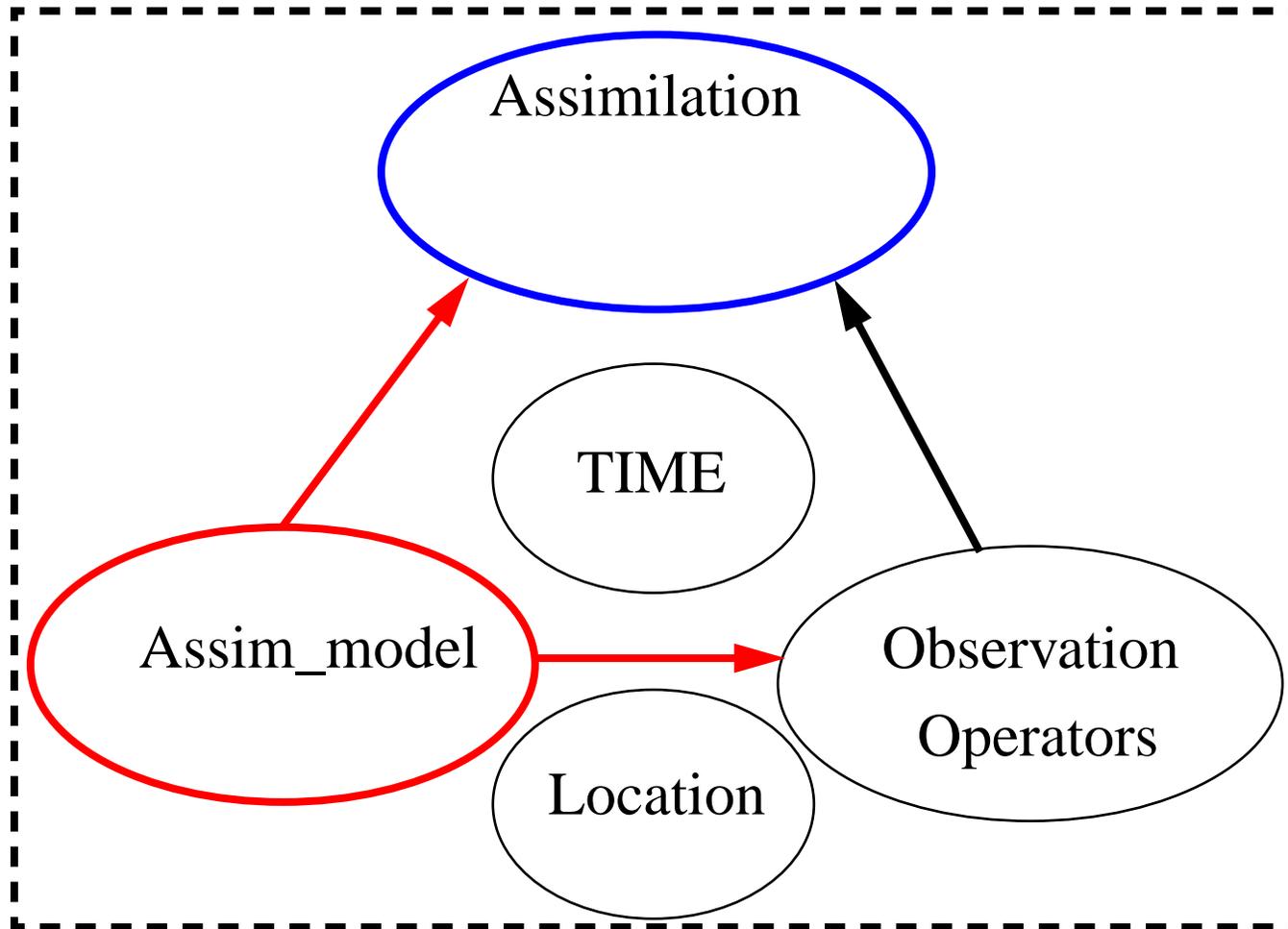
## DART compliant models

1. Lorenz-63 (3-variable model)
2. Lorenz 9-variable (low-order Primitive Equation model)
3. Lorenz-96 40-variable model
4. GFDL FMS B-grid GCM (dry dynamical atmospheric core)
5. NCAR/CGD CAM 2.0 (global climate change model)
6. WRF (NCEP/NCAR... mesoscale prediction model)
7. GFDL MOM-4 global ocean model
8. NCEP MRF in partial implementation

# Nearly illegible schematic of DART



# Separation of major components in DART core



## A Hierarchy of low-order, simple models in DART

### 1. Lorenz-63 (3-variables)

The model's 3 equations are:

$$\dot{x} = -\sigma x + \sigma y \quad (1)$$

$$\dot{y} = -xz + rx - y \quad (2)$$

$$\dot{z} = xy - bz \quad (3)$$

where the dot represents a derivative with respect to time.

Locally nearly 2D attractor manifold in 3D space

Interesting bifurcations

## What to learn with the L63 model

1. Basic DART structure
2. Design of observation sequences
3. Systematic error, filter divergence, and how to deal with them
4. Impacts of varying observation quality and frequency

## 2. Lorenz 9-variable model (9var)

$$\dot{X}_i = U_j U_k + V_j V_k - v_0 a_i X_i + Y_i + a_i z_i \quad (1)$$

$$\dot{Y}_i = U_j Y_k + Y_j V_k - X_i - v_0 a_i Y_i \quad (2)$$

$$\dot{z}_i = U_j (z_k - h_k) + (z_j - h_j) V_k - g_0 X_i - K_0 a_i z_i + F_i \quad (3)$$

$$U_i = -b_j x_i + c y_i \quad (4)$$

$$V_i = -b_k x_i - c y_i \quad (5)$$

$$X_i = -a_i x_i \quad (6)$$

$$Y_i = -a_i y_i \quad (7)$$

Defined for cyclic permutations of the indices (i, j, k) over (1, 2, 3)

X, Y and z variables ~ divergence, vorticity and height

Subscripts are zonal mean plus two wave components

## 2. Lorenz 9var model (cont.)

Has 'slow'-manifold with low gravity wave amplitude

Off-attractor perturbations can produce high-amplitude transient gravity waves

Attractor very similar to L63, quasi-2D but in 9D space

## What to learn with the 9var model

1. Diagnosing model systematic error
2. Model balance conditions and violations thereof
3. Benefits of adding noise to assimilation priors

### 3. The Lorenz-96 40-variable model

$$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$$

where  $i = 1, \dots, 40$  with cyclic indices

Low-order analog for mid-latitude tropospheric flow

## What to learn with the L96 model:

1. Dealing with ensembles smaller than model state size
2. Localization of observation impact in filters
3. Comparison of filter and variational methods
4. Dealing with large model bias
5. Comparing different varieties of ensemble filters
6. Inhomogeneous observation density in space
7. Estimating model parameters with data assimilation

## 4. GFDL's B-Grid Dynamical Core

Run-time variable resolution global PE dry dynamical core

Can be run at very low resolution

60 longitudes, 30 latitudes, 5 levels

Still produces baroclinic instability

Allows students to complete meaningful global assimilations overnight

## What to learn with B-grid core

1. Atmosphere on sphere
2. Spatially inhomogeneous model dynamics and error
3. Assess value of different types of observations
4. Spatial and temporal observation density impacts
5. How quickly things get expensive, even with only 28,800 state variables

## Computer accounts and logistics

1. MMM temporary accounts for each registered participant
2. Login 'names' are mmm01, mmm02, ..., mmm28
3. All passwords are: asp.symp

Passwords cannot be changed easily

Please be CAREFUL not to login to the wrong number account

4. Students will work in two member teams  
Limited resources  
Encourage interaction

**SIGN-UP SHEET FOR TEAMS AVAILABLE AFTER TALK**

## Computer accounts and logistics (cont.)

1. Access to 14 DEC workstations in MMM Computer lab  
Third floor of Foothills 3, look for signs  
Machines are service01, service02, ..., service14  
Initially, please use machine corresponding to your team #
2. Storage on DEC's is very limited: PLEASE BE CAREFUL  
Exercise summary document lists resource limits
3. What to do with the DEC's  
Web-browsing; WRF 3D-Var exercise set  
NOTE: type netscape & from a window, don't use menu  
4D-Var exercise set Fortran compiling, linking, executing  
Graphical analysis using Matlab, Ncview  
Mail via elm

## Computer Accounts and Logistics (cont.)

Also have accounts on compute server, ocotillo

Account names and passwords same as on DEC's

A set of 10 dual processor slave nodes (only internal network)

Single dual processor master node (on MMM network)

Slave nodes to be used for most compute tasks

telnet or ssh to ocotillo (takes you to master node)

ssh node1, or ssh node2, or ... ssh node10

Please use the slave node assigned to your group on signup

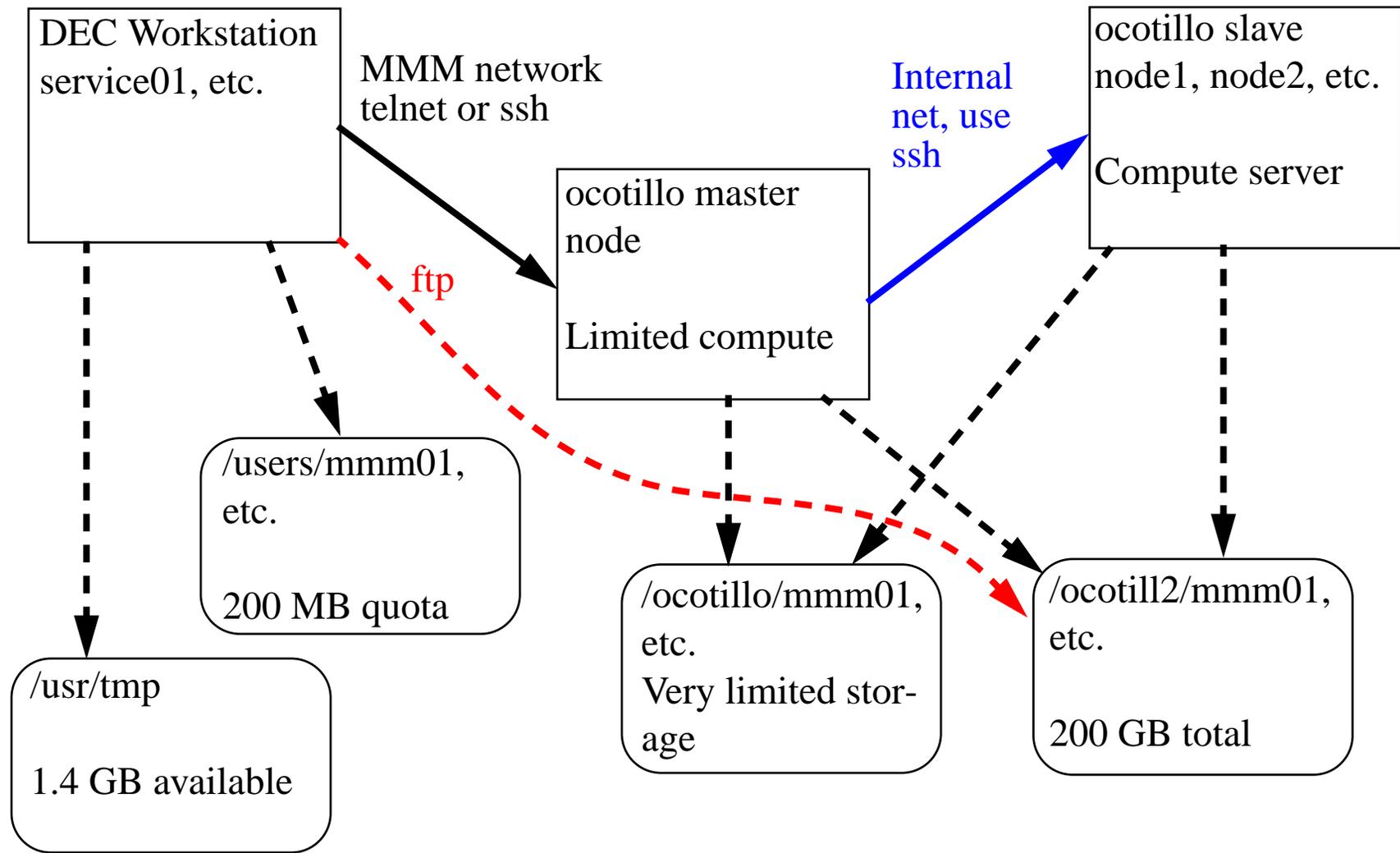
Use storage on disk /ocotillo2/mmm??

Storage limits are larger, but still be careful (DA is a hog)

Unfortunately, must ftp results from ocotillo slave nodes to DEC's

(Sorry, persistent MMM disk problems prevent cross-mount)

# Colloquium compute environment



## Suggested computing approach for DART; DEC terminal windows

Window 1:

host is ocotillo slave node, node1

Working disk is /ocotillo2/mmm01/...

DART compiling, linking, executing

Window 2:

host is DEC, service01

Working disk is /users/mmm01/...

Maintain open ftp session to ocotillo master node with target directory /ocotillo2/mmm01/...

Window 3:

host is DEC, service01

Working disk is /users/mmm01/...

Run Matlab, Ncview, or other graphical diagnostic programs here

## Getting and compiling DART code:

A basic DART directory tree is in /ocotillo2/jla/DART

Just copy from there to your working directory on /ocotillo2/

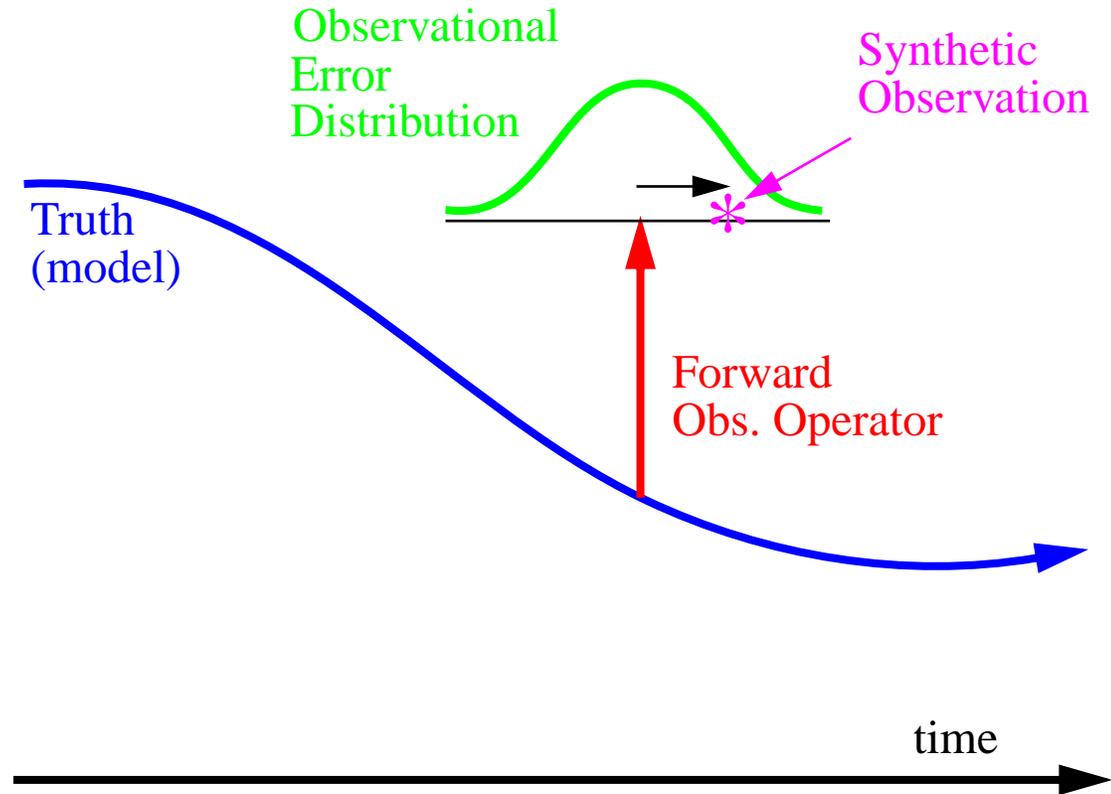
Tools for compiling and linking included, see detailed documentation in DART exercise handout

# For ASP colloquium, only ‘perfect model’ data assimilation

‘Synthetic’ obs. by applying  
‘forward observation’ operator  
to truth

(Here, this is just interpolating  
to a random horizontal location)

Instrument error simulated by  
adding random draw from a  
specified Gaussian distribu-  
tion to the interpolated obser-  
vation



Assimilation algorithm only sees these simulated observations

Result of assimilation can be compared to ‘truth’

## Four main programs support this capability for DART

Programs can be created for any of the models available

1. **create\_obs\_set\_def**: creates time-independent part of observation sequence

Contains location, type, and observational error characteristics for related set of observations.

2. **create\_obs\_sequence**: adds time information to observation set definition

Output file has complete description of a time series of observations but contains no observation values.

### 3. perfect\_model\_obs: integrates model through time specified by observation sequence definition

Operates on model state to produce synthetic observations

Produces observation sequence output file (definition + value)

Produces NetCDF diagnostic file, True\_State.nc

### 4. filter: assimilates observations in observation sequence file

Prior\_Diag.nc contains prior ensemble (before assimilation)  
diagnostics

Posterior\_Diag.nc has posterior (after assimilation)

## Graphical diagnostic tools for NetCDF output

1. Ncview: provides simple animations of 'slices' of data, mostly for B-grid GCM
2. Matlab: a sequence of matlab scripts provide diagnostics for all models

## Other helpful resources:

1. NCAR library is available two floors above us  
Hardcopy journals  
Copying facilities
2. Web-access to NCAR library on-line journals  
Go to [www.ucar.edu](http://www.ucar.edu) and search to get to library
3. Web-access to AMS journals on-line  
[www.ametsoc.org](http://www.ametsoc.org)
4. List of web-pages with relevant DA stuff  
Will be available off DART web-site this afternoon

## Student Participation:

1. Please participate in discussions!!!
2. Please keep a journal describing your experience with the exercises and provide us with a copy at the end of the colloquium
3. Some students have volunteered to make short presentations about their ongoing research
4. Students are welcome to make short presentations about any relevant topic during the discussion on Friday