Parallel Implementation of Ensemble Filter Algorithms for Data Assimilation

Nancy Collins
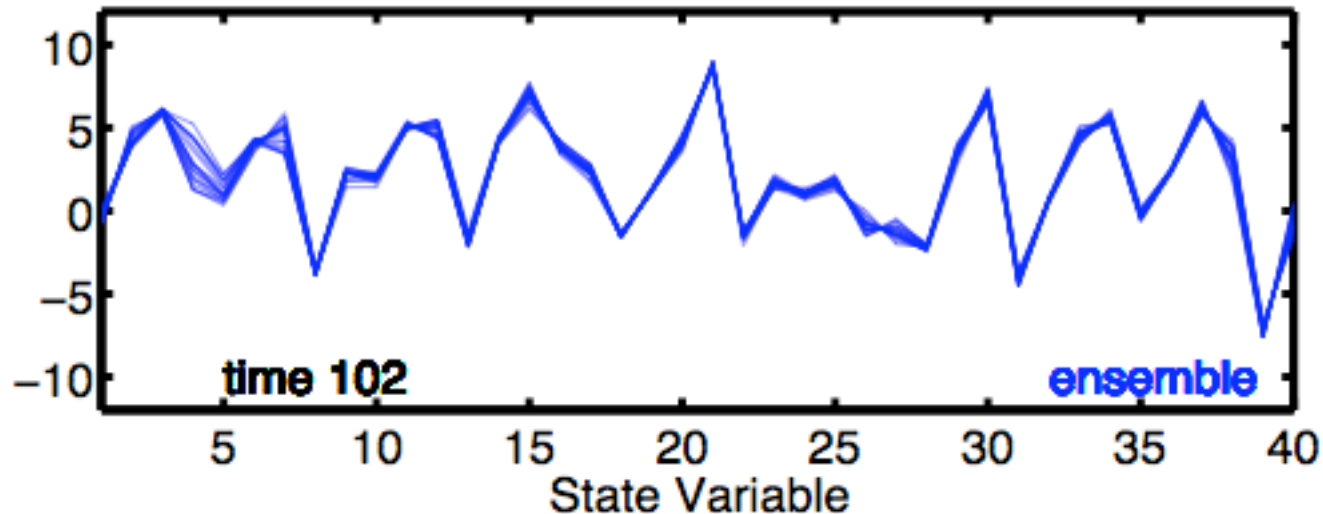
NCAR/IMAGe

Supercomputing 2007

Outline

- Data Assimilation Overview

- DART software
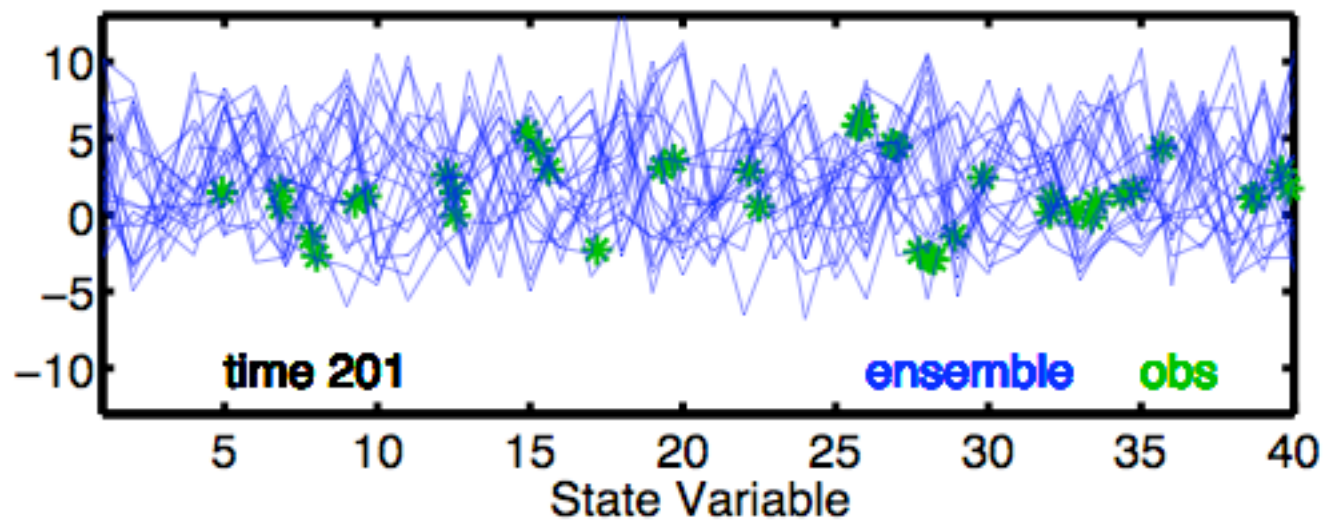
- Examples

- Parallel Algorithms

- Summary

What is Data Assimilation?

- Given:
  - A real physical system (e.g. atmosphere)
  - Observations of that system
  - A computer model of the system
- You can get:
  - Better estimates of the state of the system
  - Better estimates of the observation errors
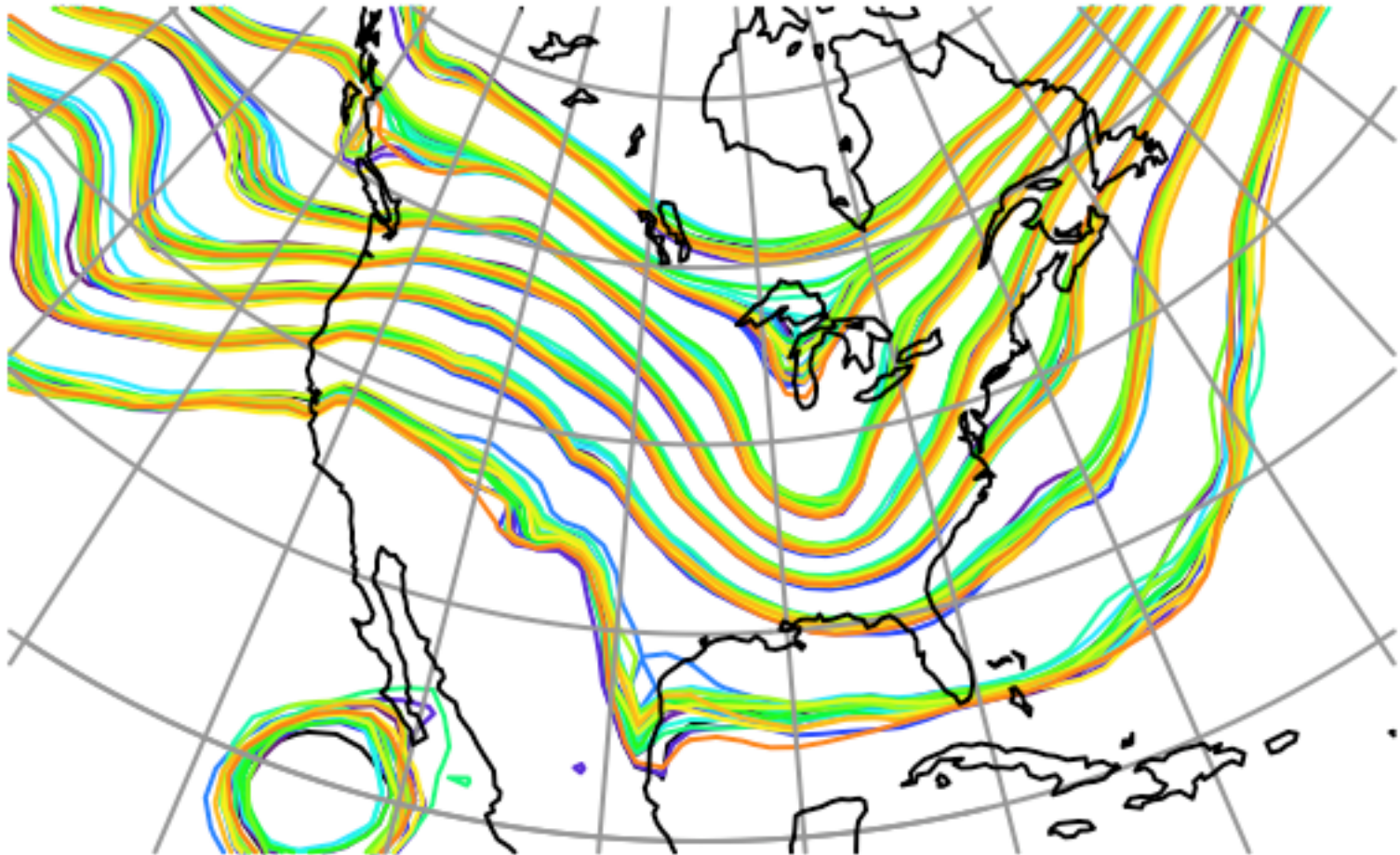 - A better model of the system

Simple 1D (chaotic) model

1D model with Observations

Seeing Model Uncertainty

The DART project

- The Data Assimilation Research Testbed (DART) is a community software toolkit

- A collection of software tools to explore DA with various models and observations.

- It must be general and yet deal with high stress computational requirements
  - Running many ensembles (copies of a model)
  - Large datasets/models

DART features

- A working Data Assimilation system
- Observation support
  - Observations are handled independently of which model is being used
  - Code handles existing NCEP observation types
  - Support to add novel observation types (e.g. GPS occultation) with a minimum of coding
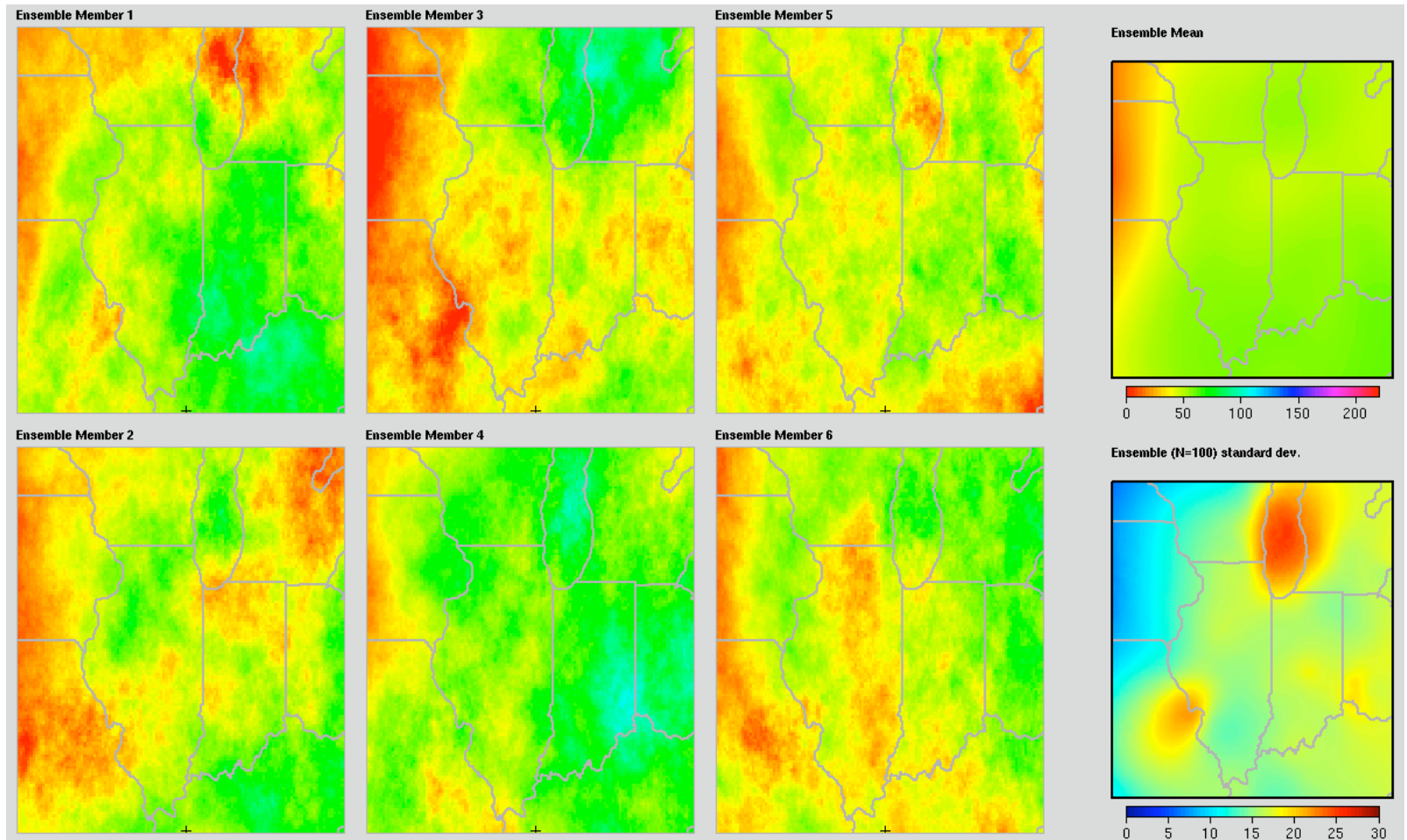
DART features (cont)

- Model support
  - Includes working models - from 1D simple models to complete atmospheric models
  - Code to add new models without being very intrusive or needing to know the model internals
- Tutorial materials for learning and teaching DA
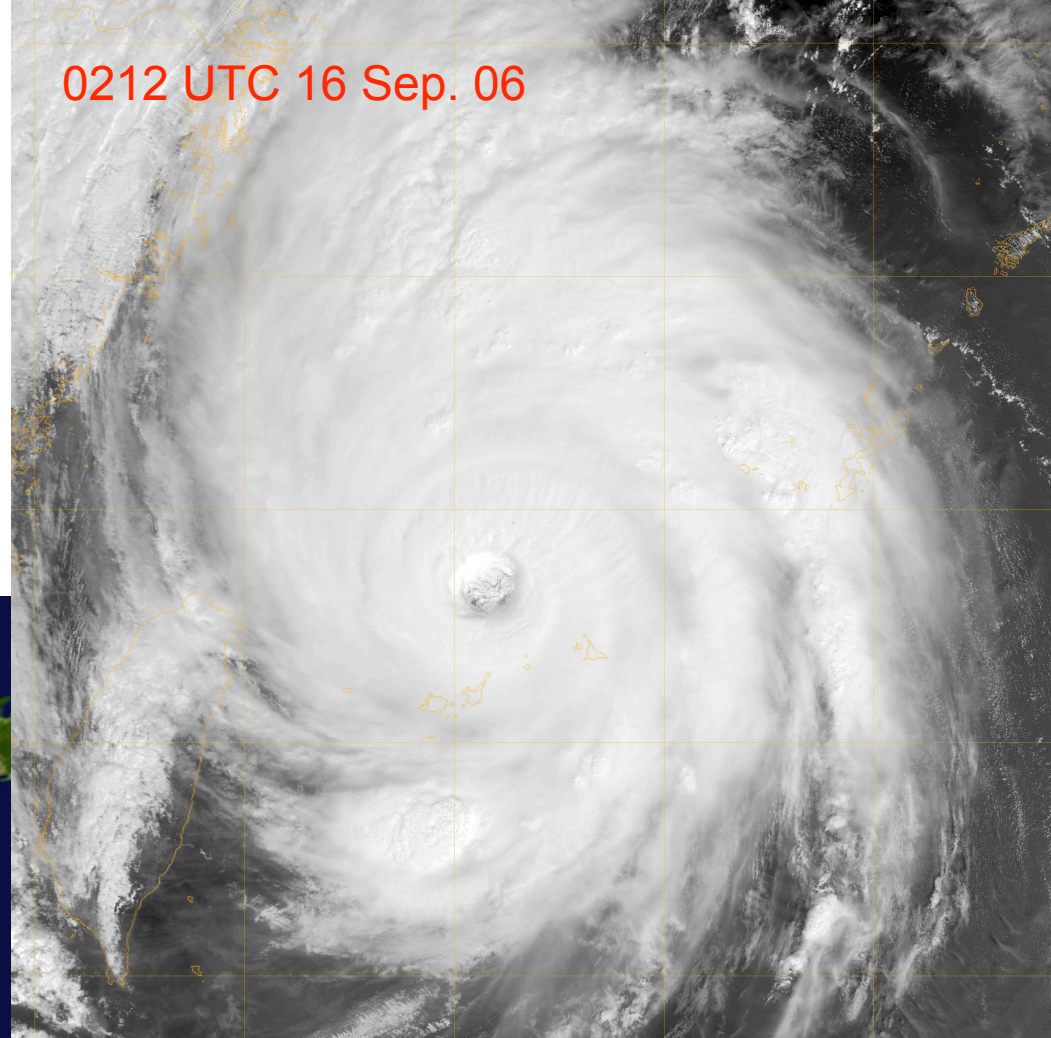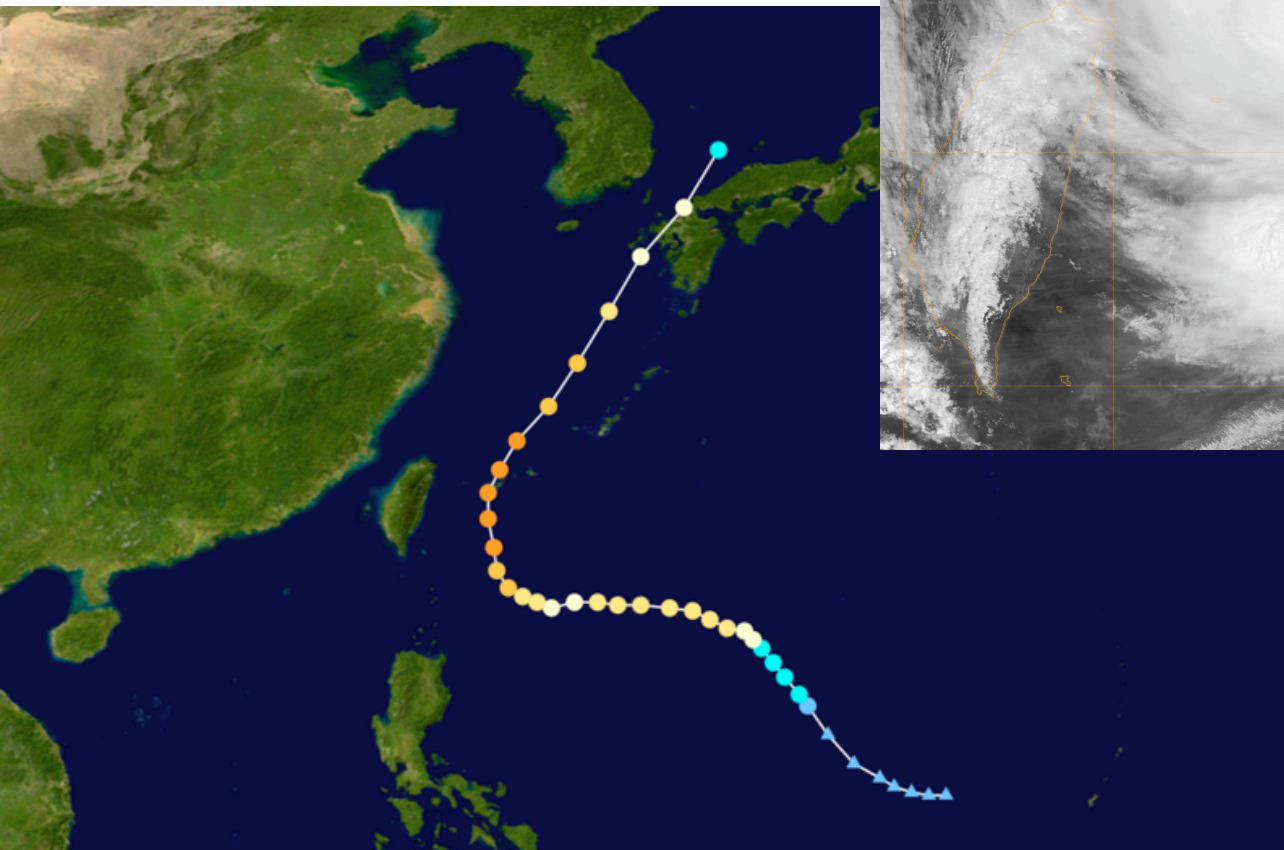- Code to do research on novel DA algorithms

DART uses

- Initial conditions for forecasts
- Predicting what observation types and locations will make the largest impact on a model
- Accounting for errors in the model itself
- Estimating model parameters which are unknown or difficult to understand
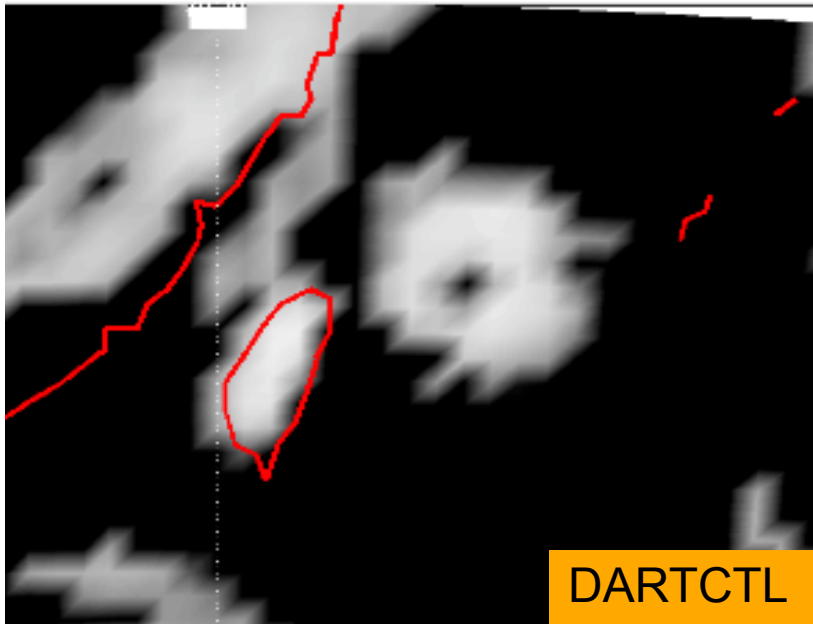
Ozone near Chicago
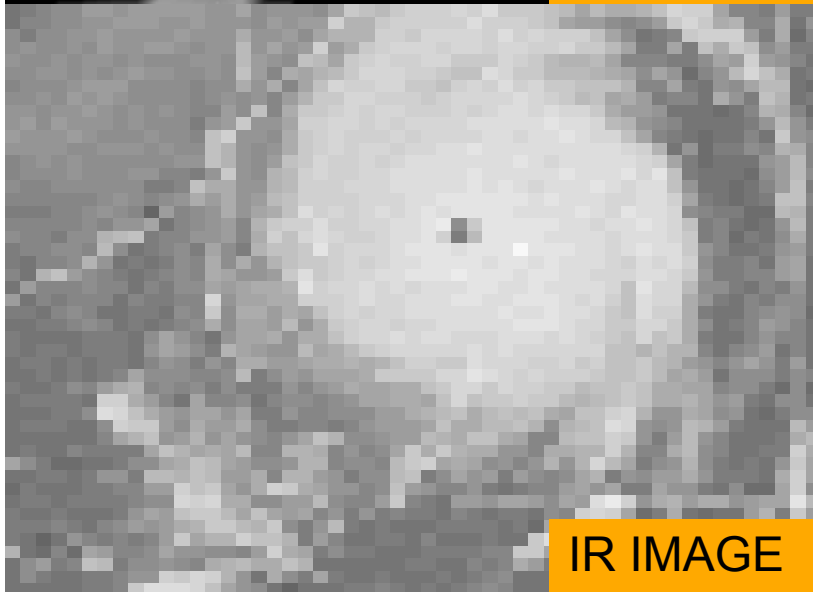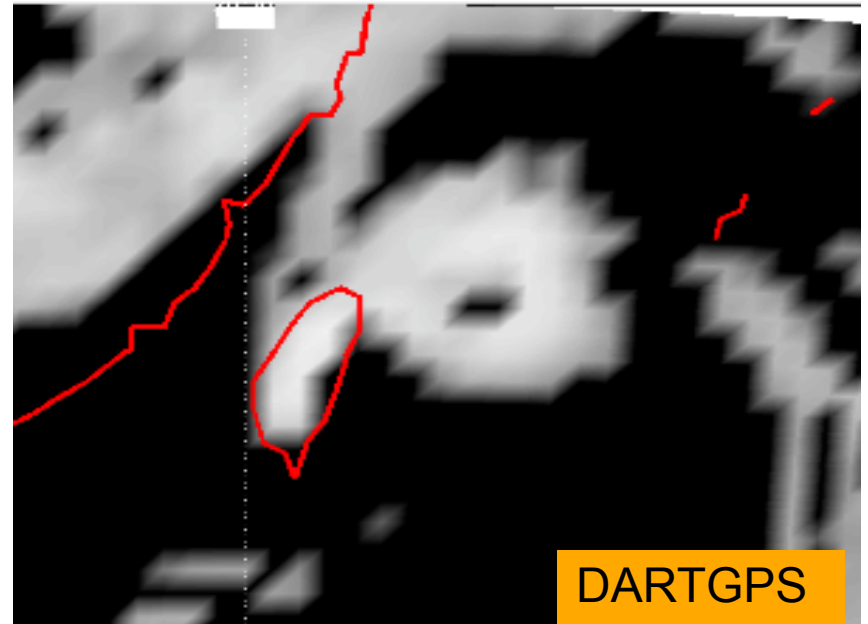
Typhoon Shan Shan 2006

0212 UTC 16 Sep. 06

Formed: 9 Sep 06

Dissipated: 19 Sep 06

Lowest pressure: 919 mb

Category 4 storm

Typhoon Shanshan - DART with GPS obs

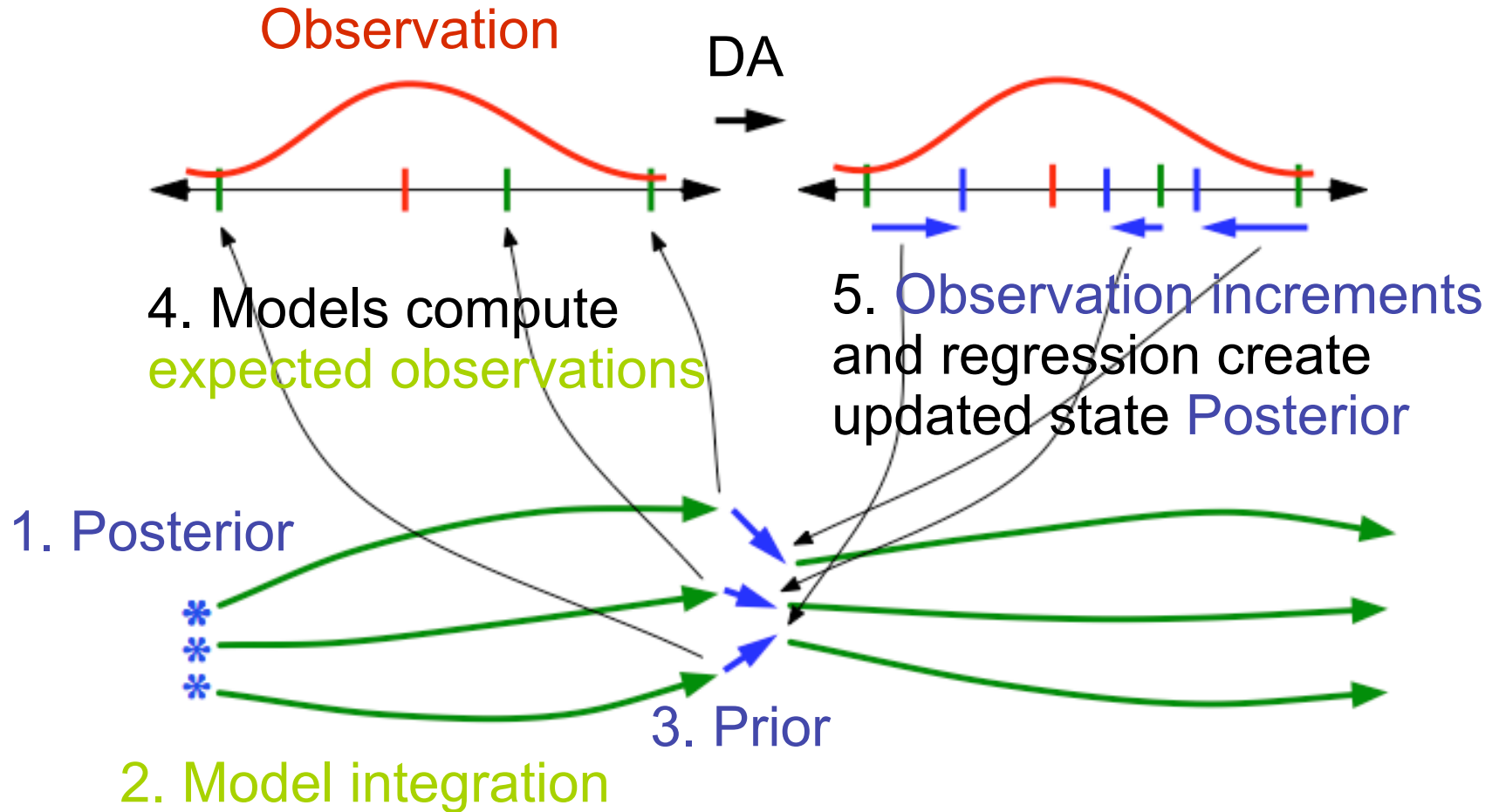DARTCTL, Total Cloud Water [log(kg/kg)], 2006-09-16_0(

DARTGPS, Total Cloud Water [log(kg/kg)], 2006-09-16_0(

DARTCTL

DARTGPS

IR IMAGE

High level view of process

Observation

DA

4. Models compute
expected observations

5. Observation increments
and regression create
updated state Posterior

1. Posterior
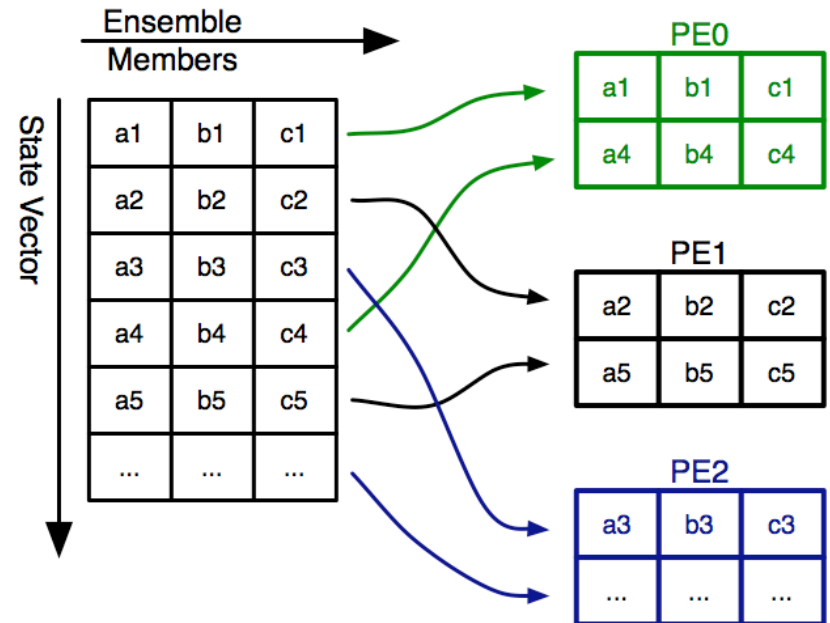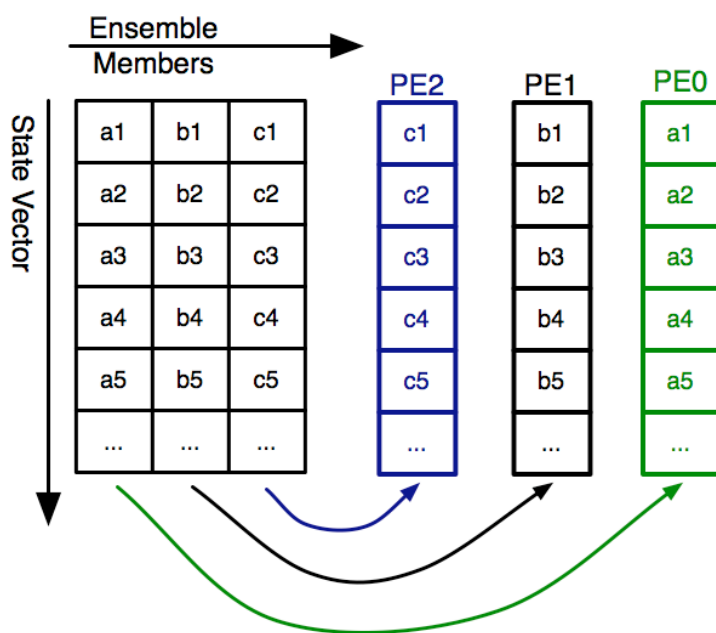
3. Prior

2. Model integration

Parallel Implementation

- Must be generic across hardware architectures and compilers (MPI)
- Must be bitwise reproducible
- Must scale reasonably well
  - Large numbers of models running at once
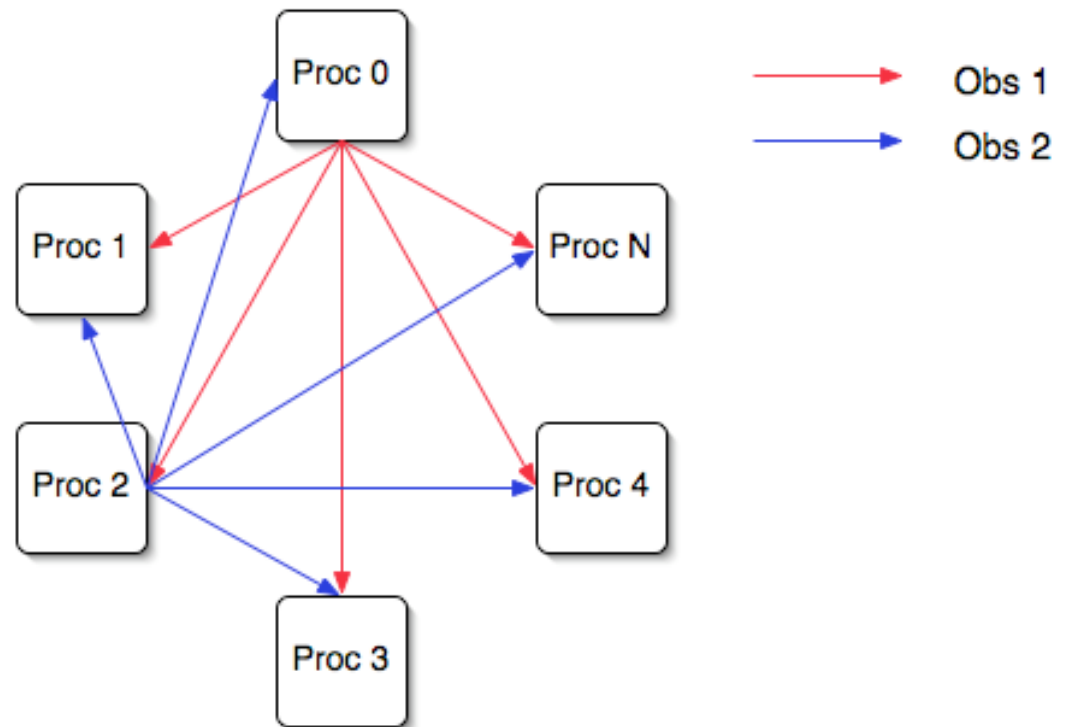 - Large models/dataset sizes

Parallel Implementation
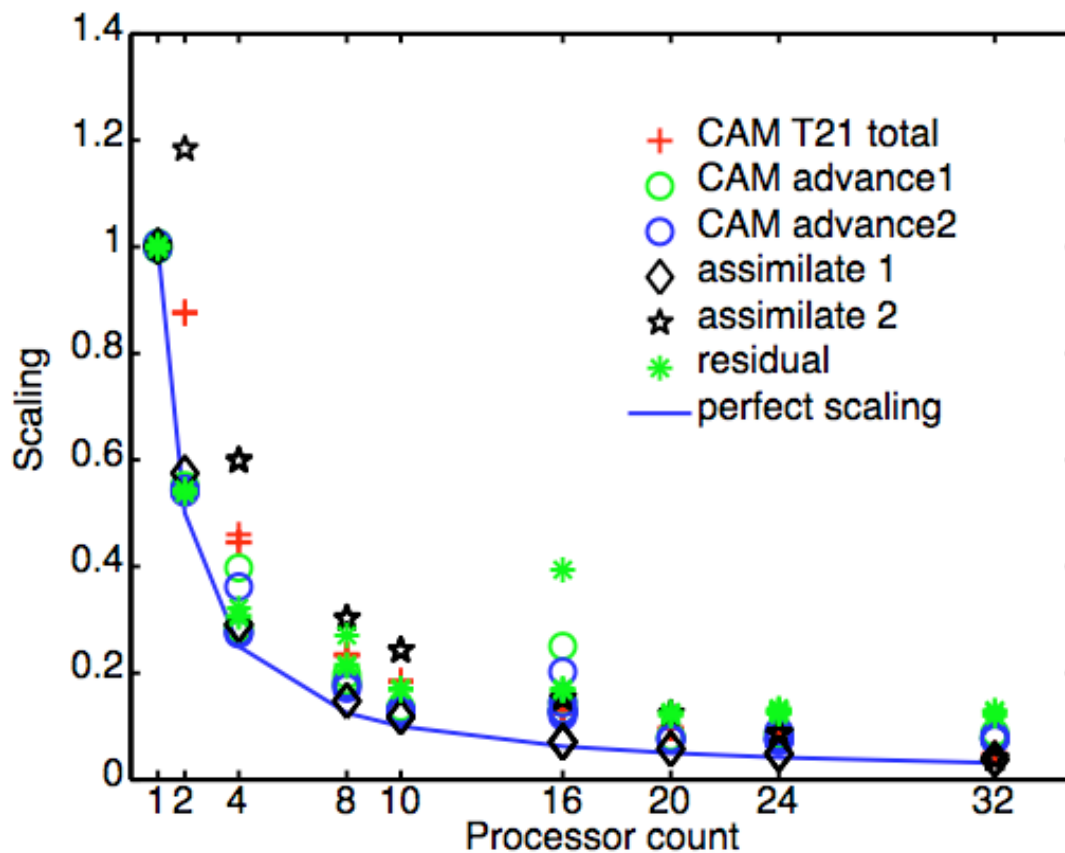
All-to-all transpose:

Parallel Implementation

Bitwise reproducibility
with the serial code

Parallel Performance

Summary

- DART is a community DA facility
- It solves interesting problems
- It has immense computing requirements
  - Multiple ensembles
  - Large models
  - Large datasets
- The algorithms in DART parallelize well

For more info

- DART download and web page:
  - http://www.image.ucar.edu/DAReS/DART/
- Email me:
  - nancy@ucar.edu
- The DART group:
  - Jeff Anderson, Tim Hoar, Nancy Collins, Kevin Raeder, Hui Liu, plus an active user community

- J.L.Anderson and N.Collins, "Scalable implementations of ensemble filtering algorithms for data assimilation", J. Atmos. Ocean. Tech., 24, Sept 2007, 1452–1463.