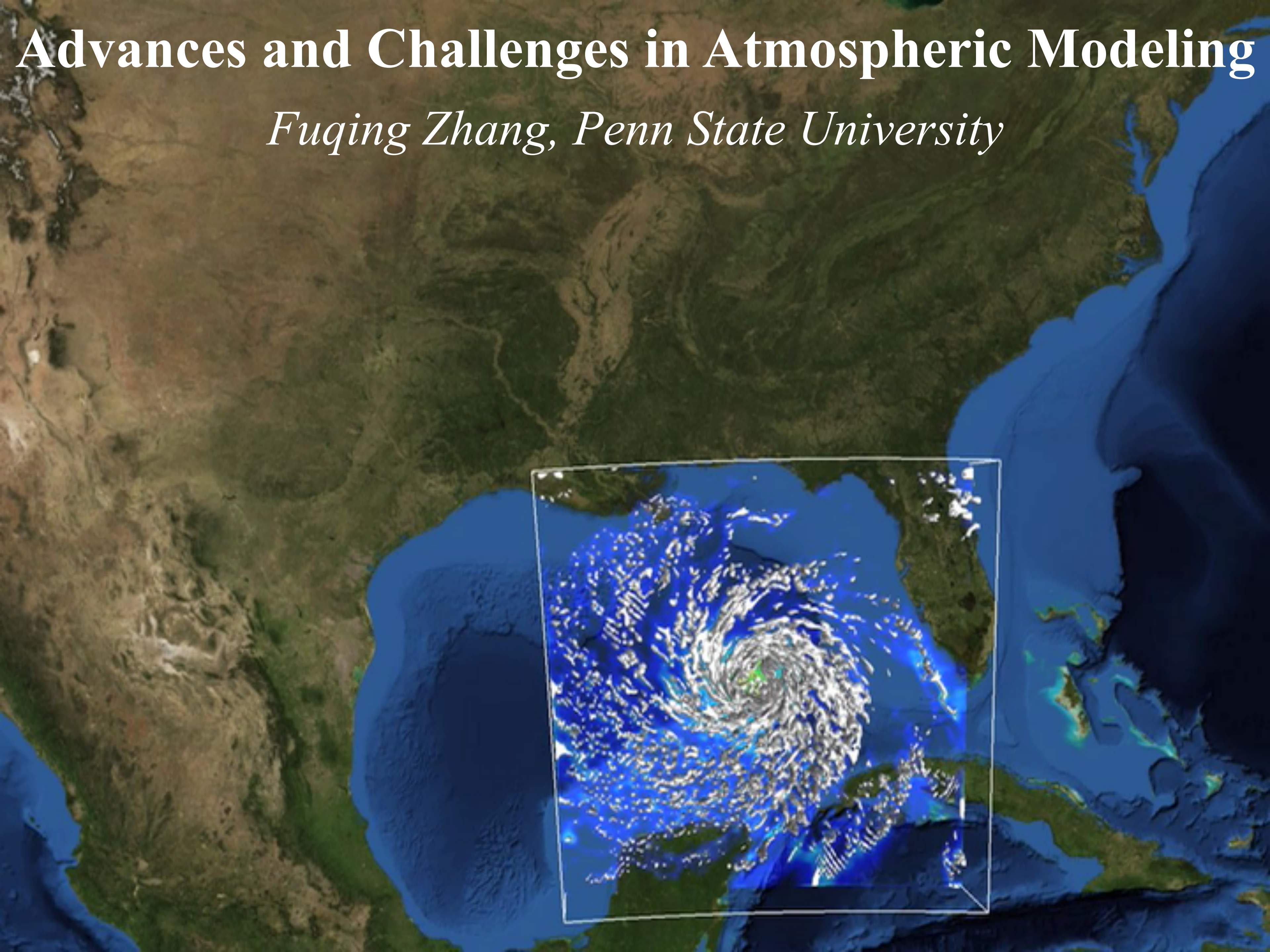
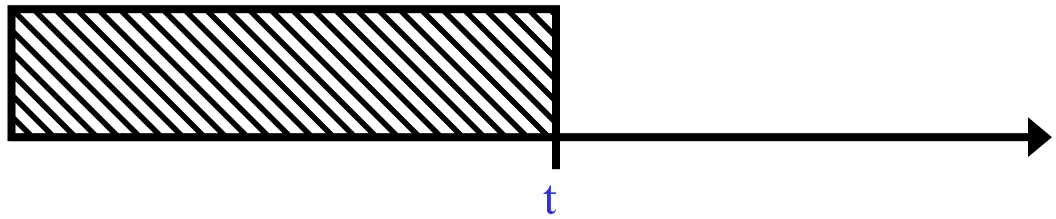
Advances and Challenges in Atmospheric Modeling

Fuqing Zhang, Penn State University

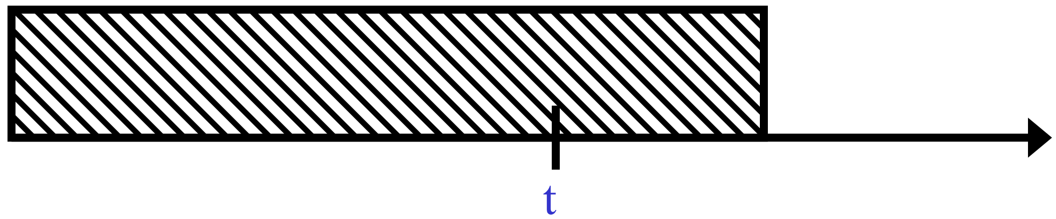


Types of State Estimation (dimensions ~10^{10})

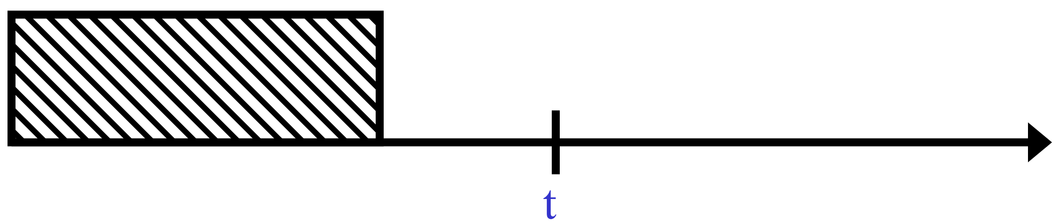
(1) Filtering: When time of desired estimate (t) coincides with the last measurement point



(2) Smoothing: When time of desired estimate (t) falls within the span of available measurement data



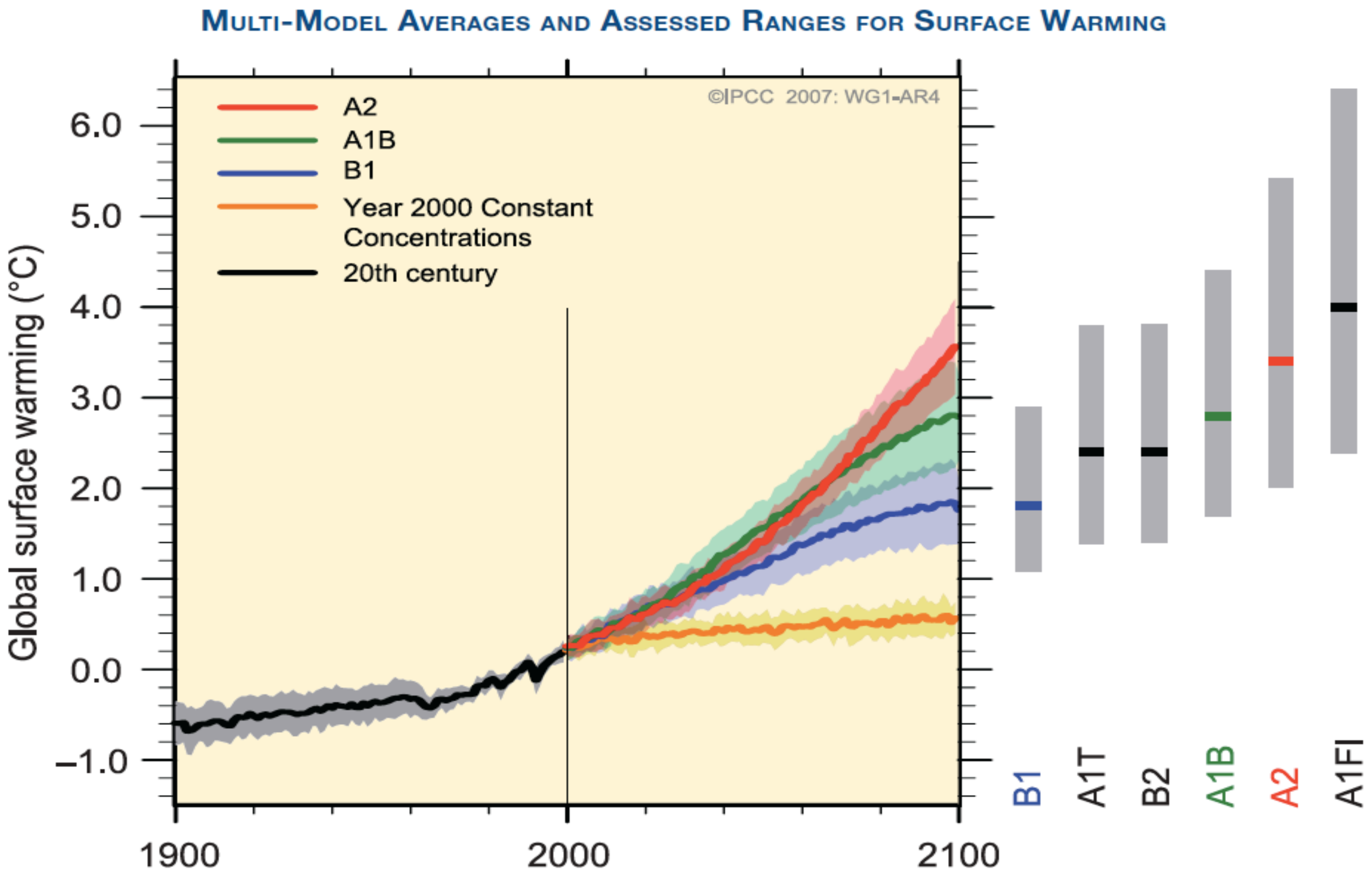
(3) Prediction: When time of desired estimate (t) occurs after the last available measurement



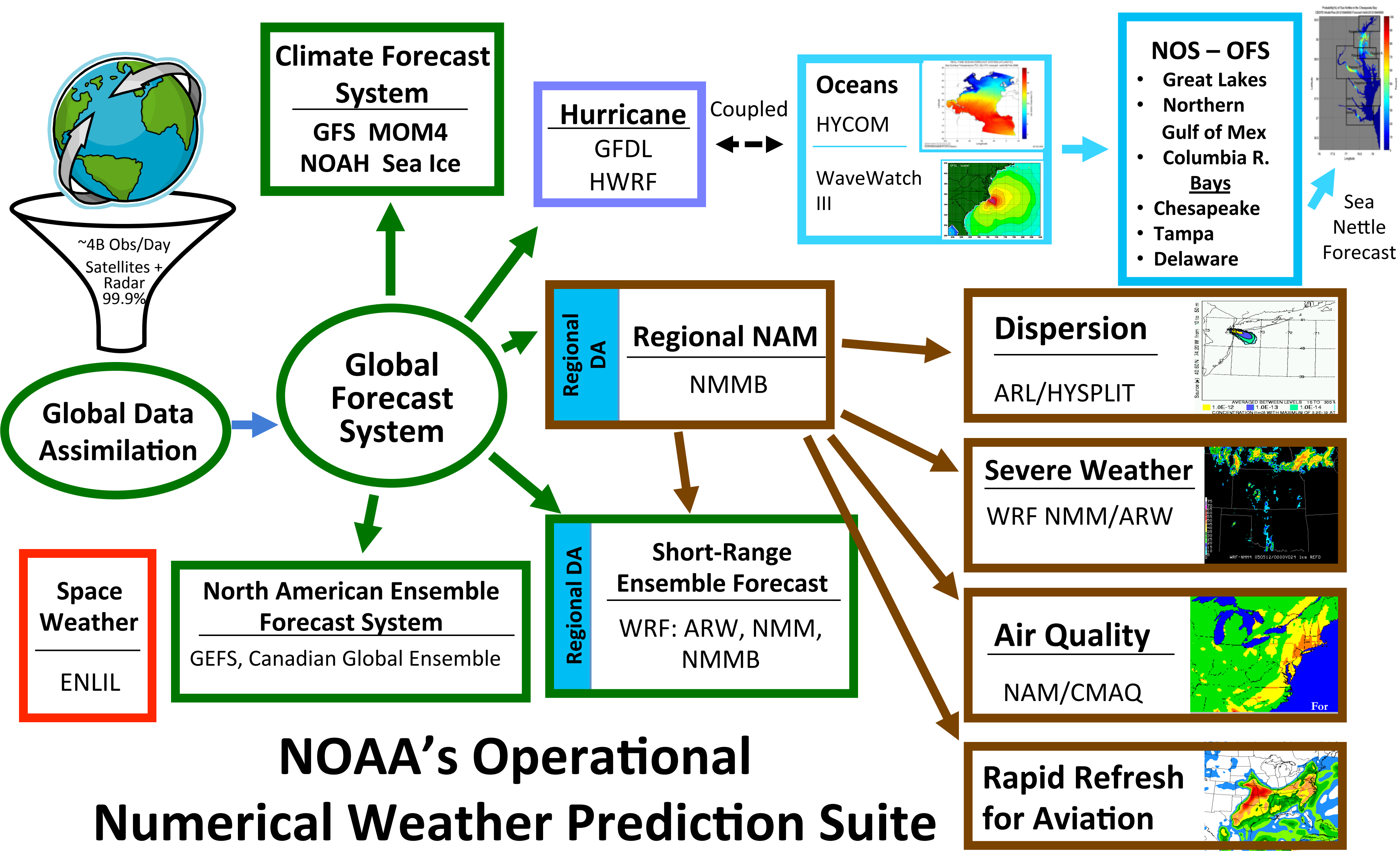
Atmospheric modeling: the process of numerically solving prognostic partial differential equations based on fundamental conversation laws, (e.g., momentum, energy, mass, moisture) that govern atmospheric fluid

Data Assimilation: the process to obtain the best estimate by combing the prior estimate (from model forecast) and new observations, along with associated uncertainties.

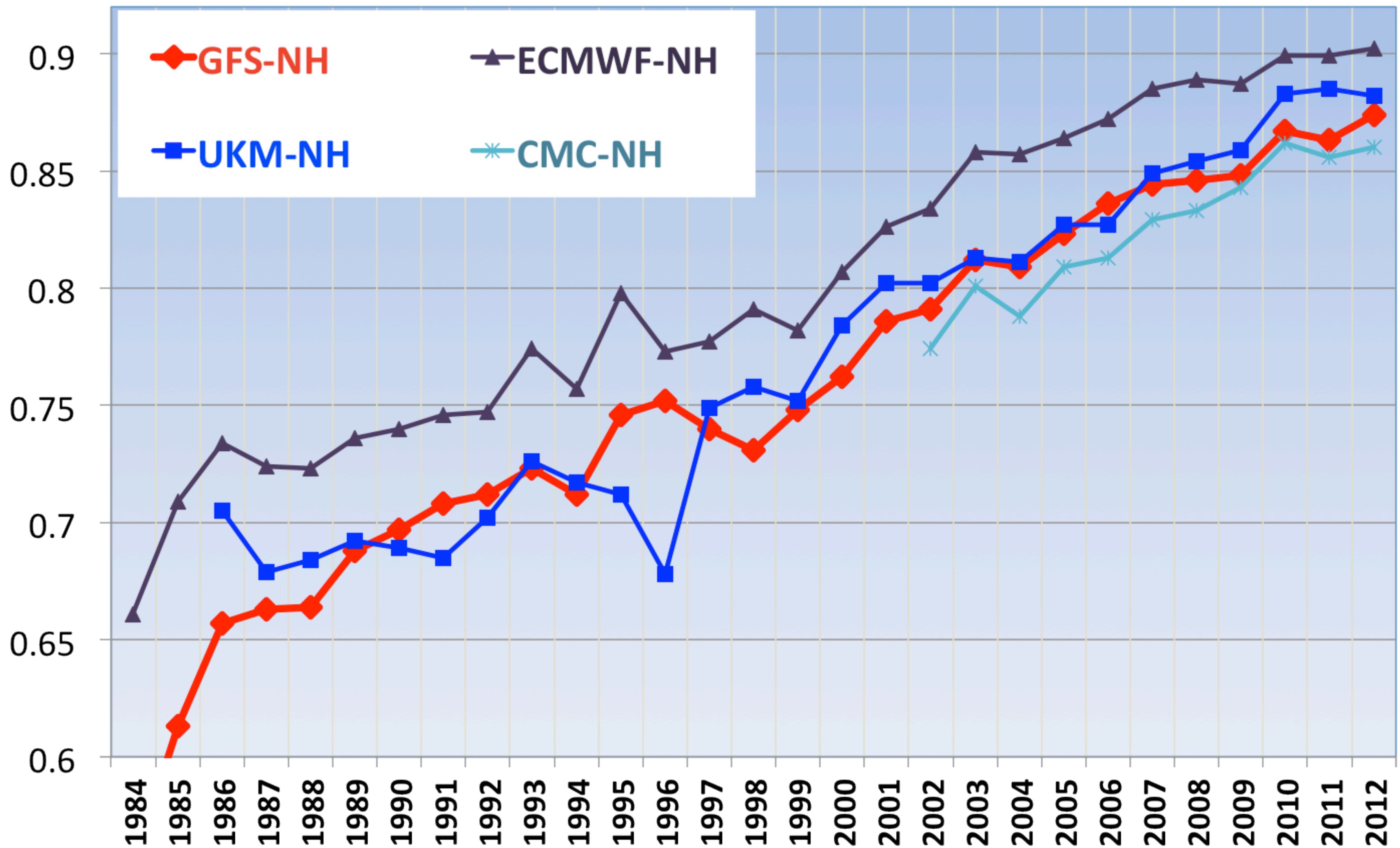
What are we modeling: Climate prejection



What are we modeling: Weather prediction

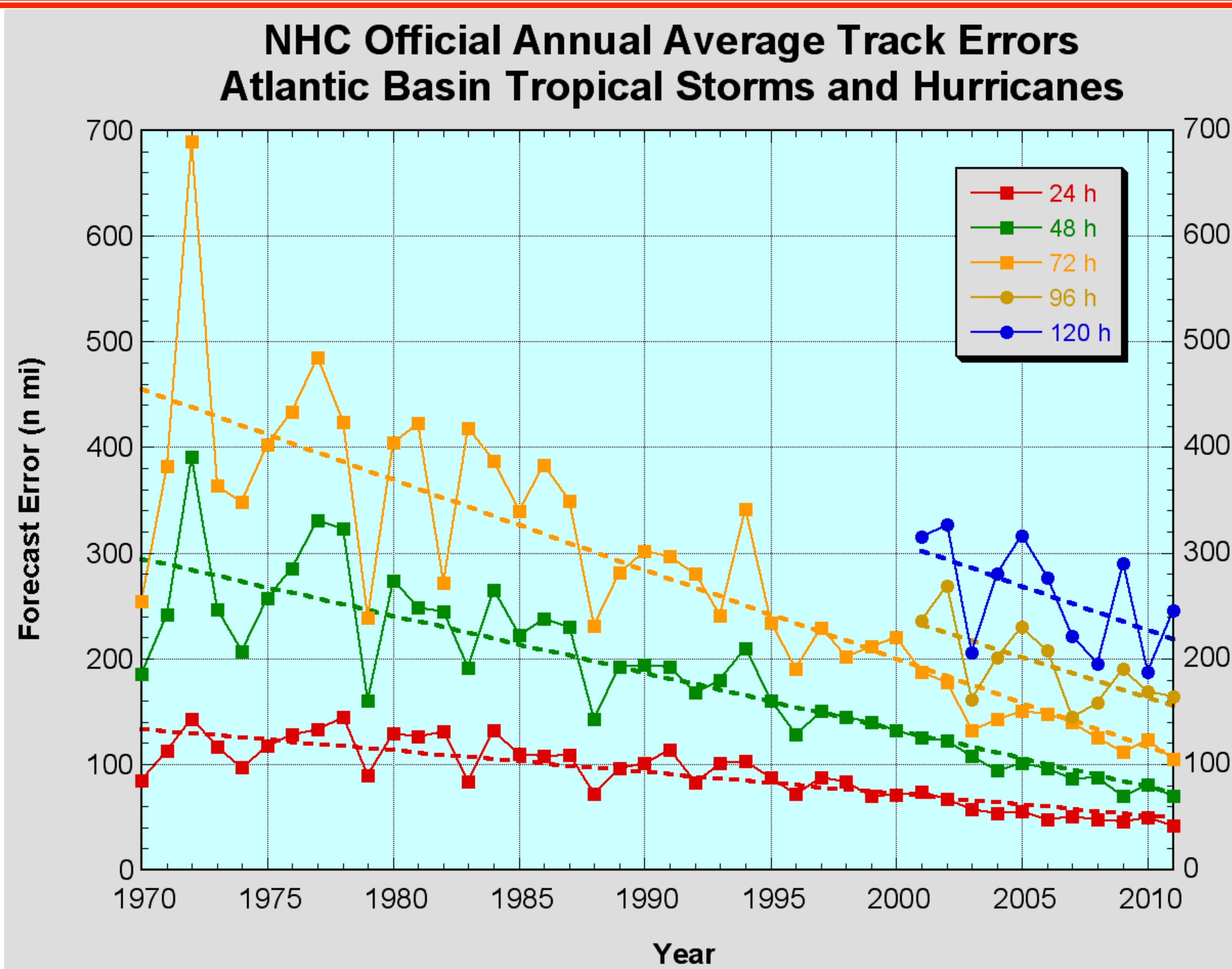


Annual Mean NH 500hPa height Day-5 Anomaly Correlations

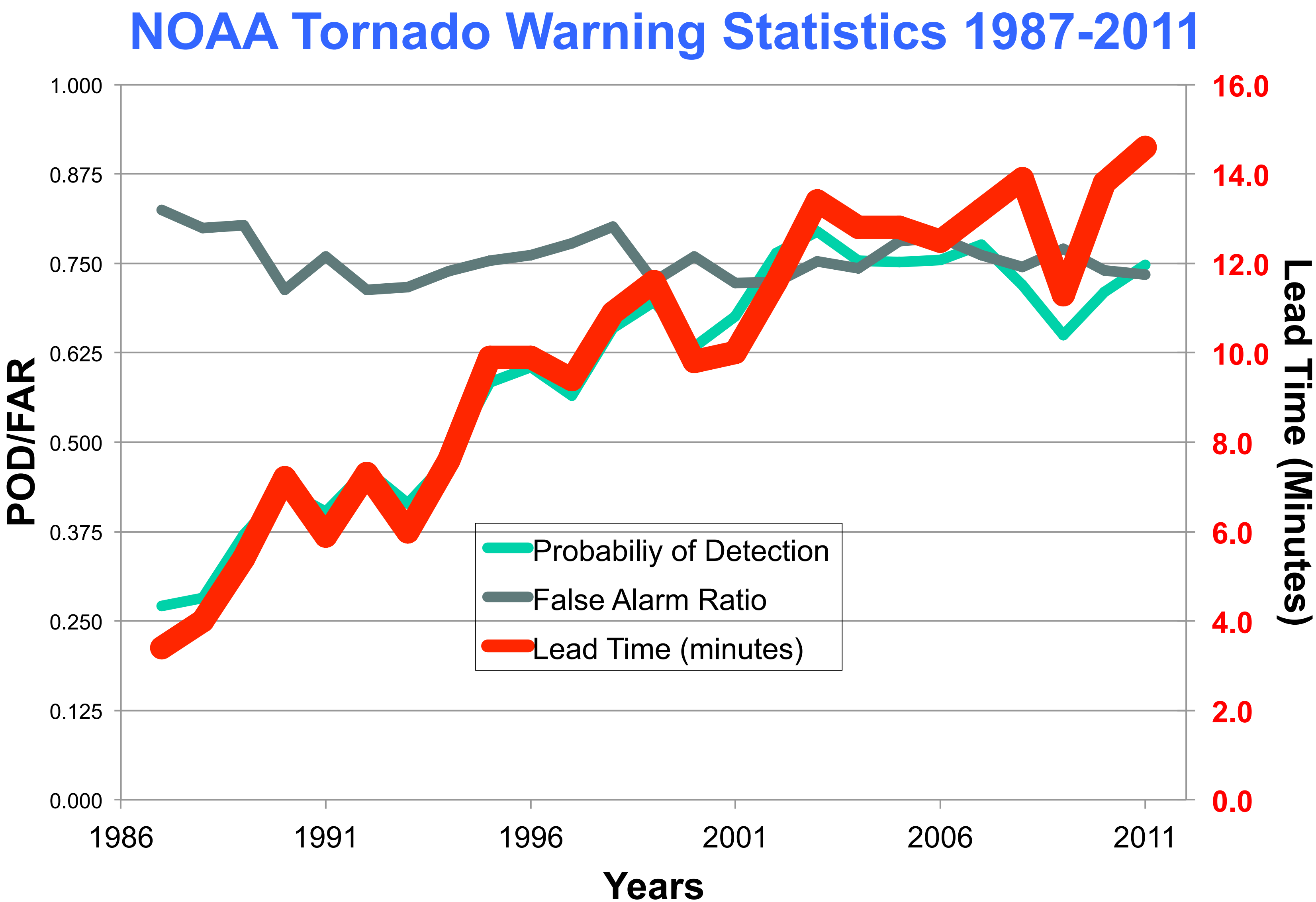


- ECMWF, GFS and CMC were better in 2012 than in 2011. GFS has the largest gain.
- UKM and FNOMC were slightly worse in 2012 than 2011.

How Far We've Come



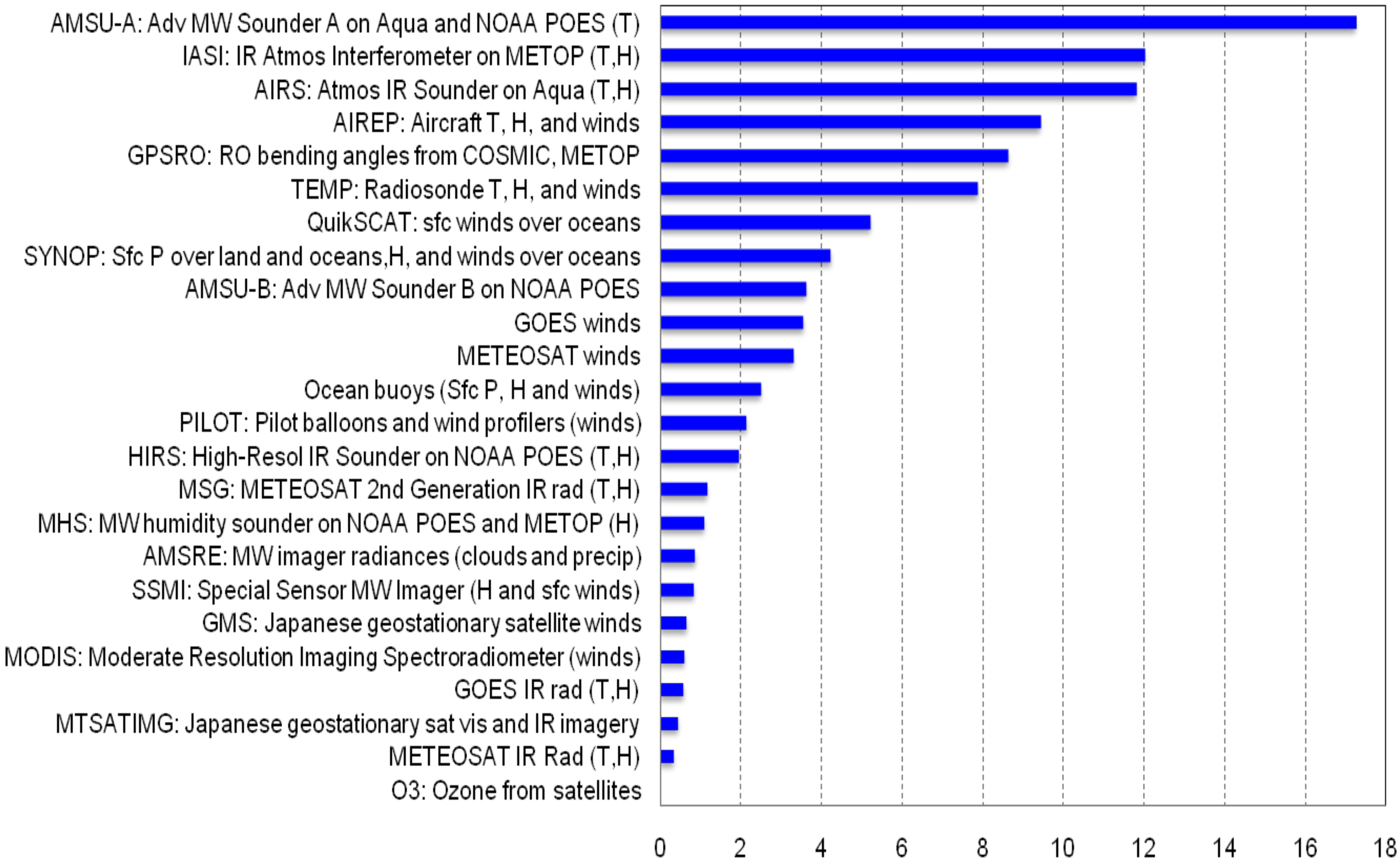
How Far We've Come



How We Got There

1. Better resolution: both horizontal & vertical, and in time
2. Better dynamics and numerics: more accurate formulations with fewer approximations / simplifications
3. Better physics: better representation of small-scale processes like radiation, clouds, precipitation & turbulent exchange of heat, moisture and momentum
4. Better data: Increase in available observations, especially from remote sensing instruments such as satellites/radars
5. Better data assimilation: more accurate analysis and 4-dimensional data assimilation producing improved initial conditions
6. Use of ensembles: for uncertainty and probabilistic estimates for the background prior, analysis posterior, and forecast
7. Better computing: increase in in data processing and computer power by ~10^15 times since 1955 of the first NWP to current

Importance of Observations and Data Assimilation: Impact of satellite data on global NWP…..



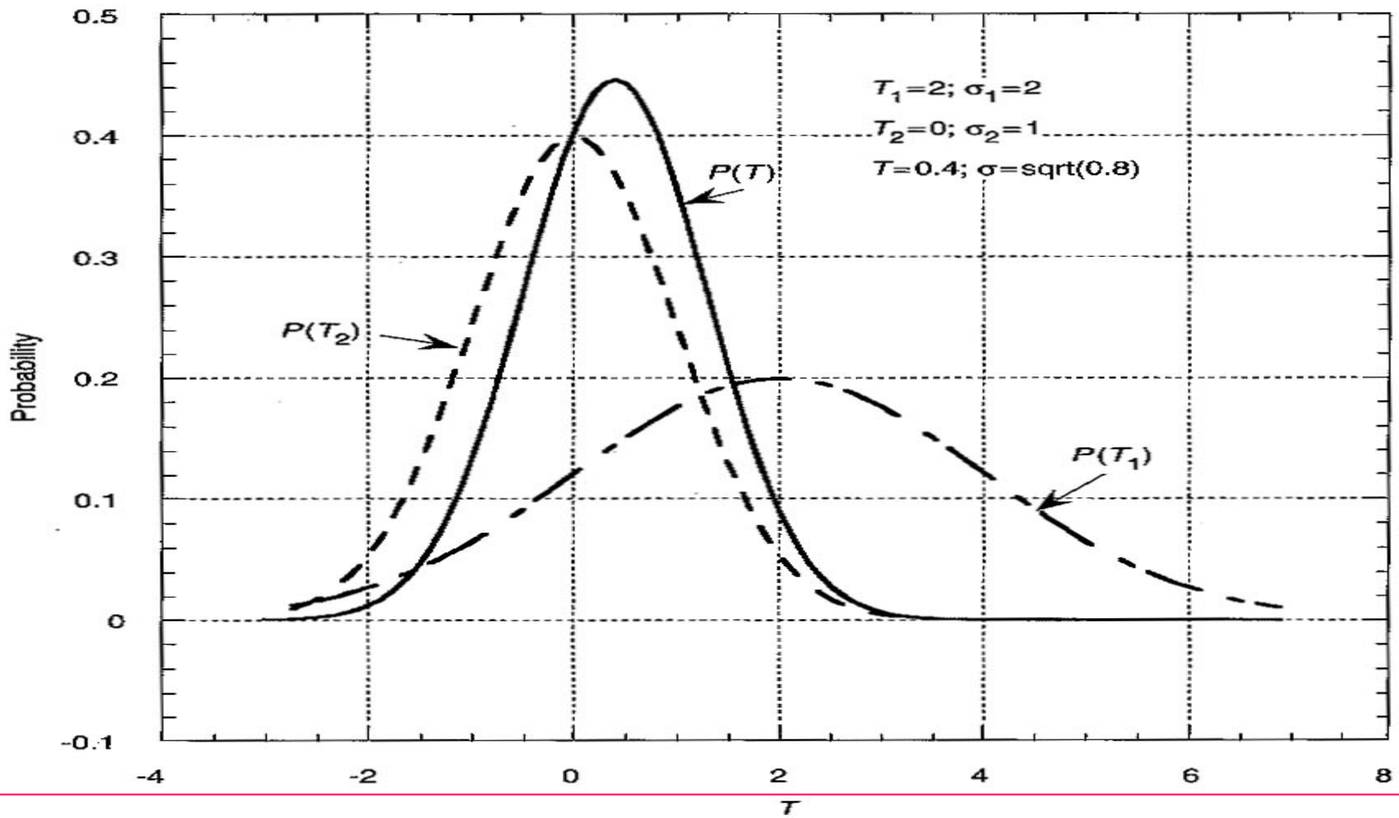
Data Assimilation: A Simple Example

If T_1 is the background state with std of σ_1 ; T_2 is the observation with std of σ_2 , then the posterior analysis T and its std σ will be:

$$T = T_1 + K\,(T_2 - T_1); \quad K = \sigma_1^{\,2}/(\sigma_1^2+\sigma_2^2)$$

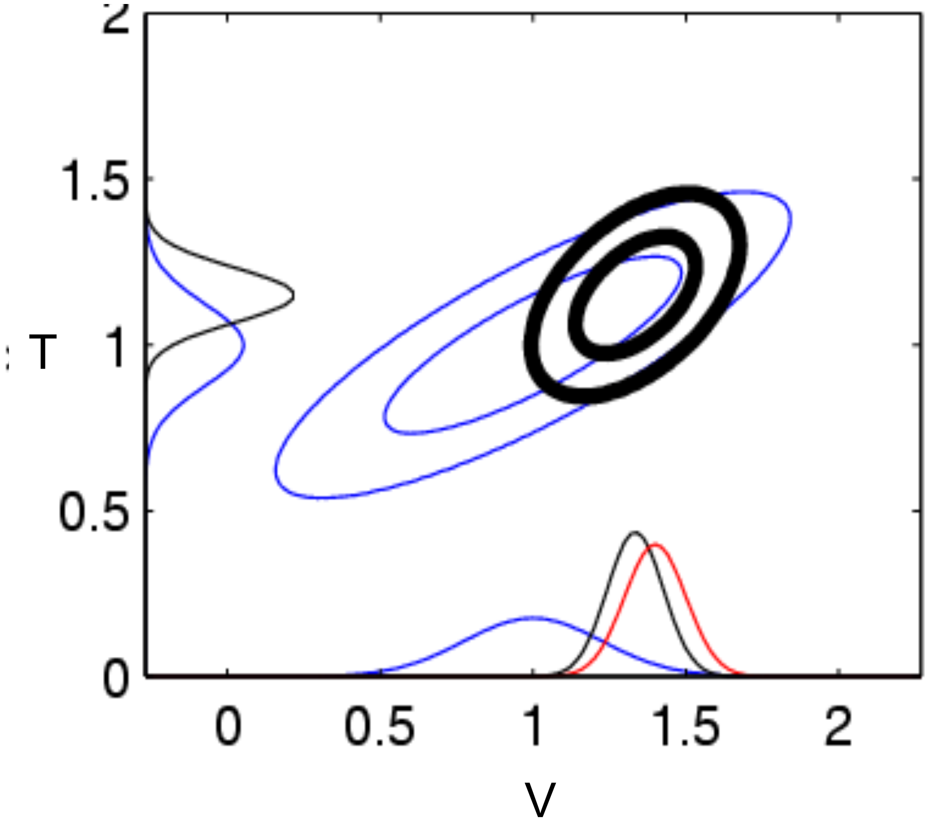
or $$T = [\sigma_2^{\,2}/(\sigma_1^2+\sigma_2^2)]\,T_1 + [\sigma_1^{\,2}/(\sigma_1^2+\sigma_2^2)]\,T_2$$

$$\sigma^2 = \sigma_1^{\,2}\,\sigma_2^2\,/(\sigma_1^2+\sigma_2^2)$$



Data Assimilation: A Multivariate Example

If T_1 is unobserved but with measurements of V_2 (e.g., radar radial velocity obs), then the posterior analysis T depends on background error covariance Cov(T,V):



$$T = T_1 + K\,(V_2 - V_1)$$

$$K = Cov(T,V)/(\sigma_{v1}^2 + \sigma_{v2}^2)$$

In EnKF, Cov(T,V) is estimated by short-range ensemble and is flow-dependent

Prevailing data assimilation techniques: 3DVAR

3D Variational assimilation (3D-Var)

The analysis is found as the model state x corresponding to the minimum of the (scalar) cost function

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x^b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x^b})}_{J_b} + \underbrace{\frac{1}{2}(H\mathbf{x}-\mathbf{y})^T\mathbf{R}^{-1}(H\mathbf{x}-\mathbf{y})}_{J_o}$$

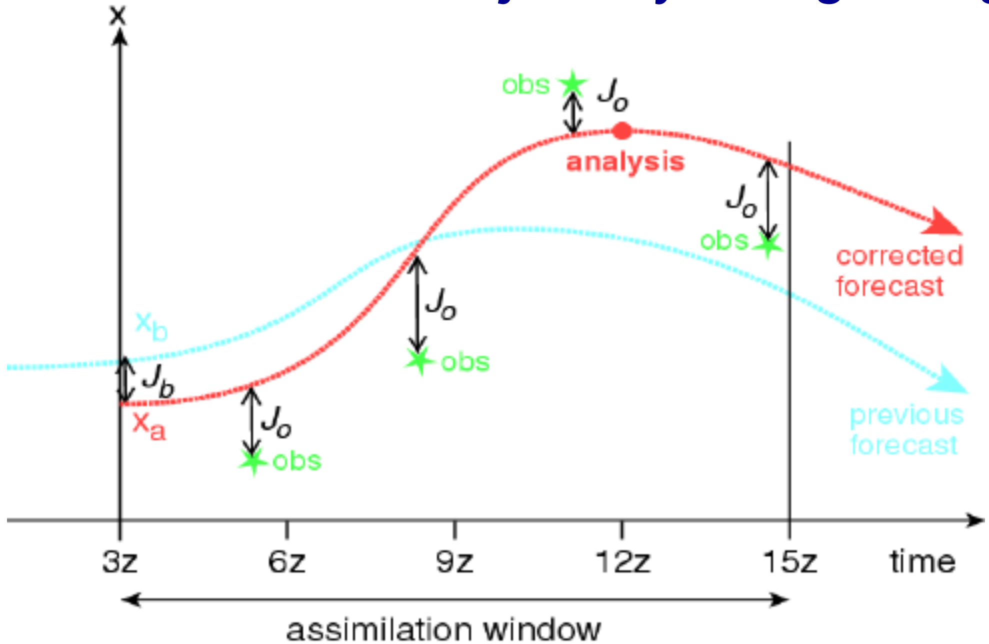
where

J_b measures the departure from the background field x^{b}

J_o measures the departure from the observations y

Prevailing data assimilation techniques: 4DVAR

Minimization across time & a trajectory fitting using a adjoint model

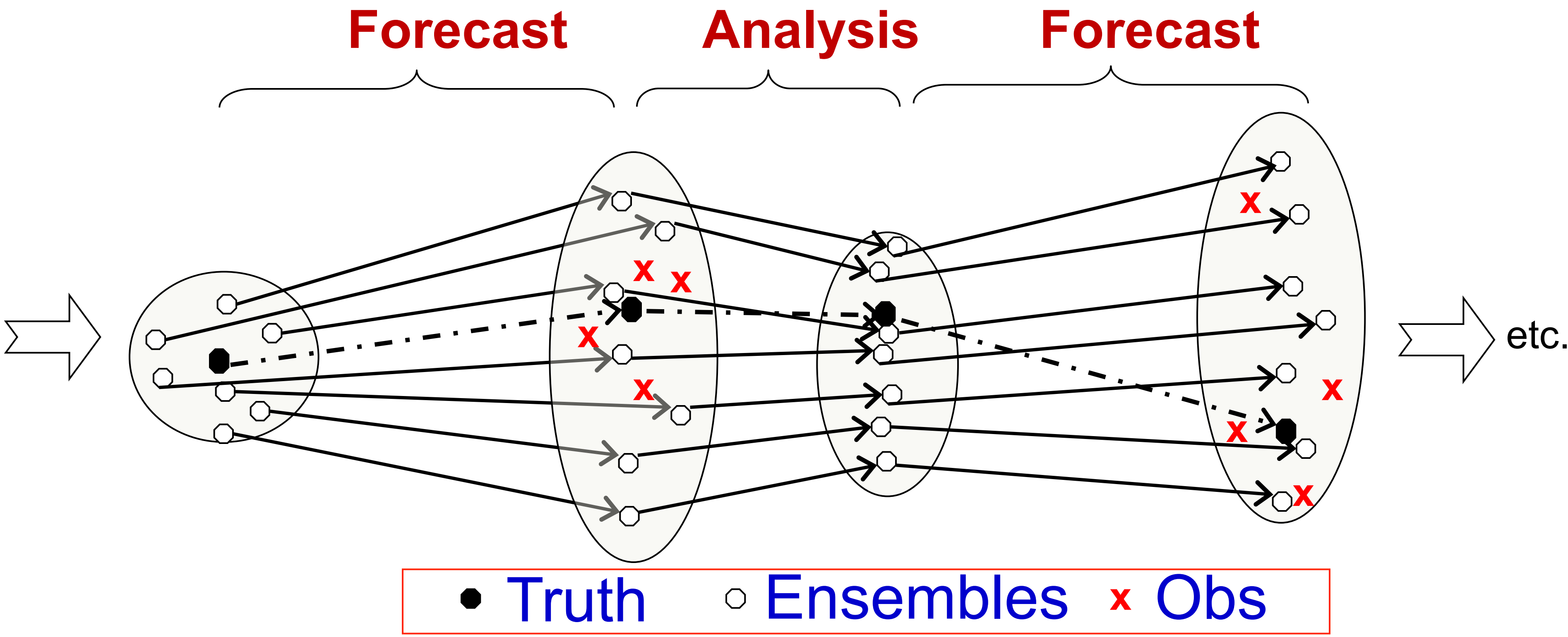


$$J = J_b + J_o \quad \begin{cases} J_b = \dfrac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}^b\right)^T \boxed{\mathbf{B}^{-1}} \left(\mathbf{x}_0 - \mathbf{x}^b\right) \\ J_o = \dfrac{1}{2}\sum_k \left[\mathbf{y}_k - H\left(M_k\left(\mathbf{x}_0\right)\right)\right]^T \mathbf{R}^{-1}\left[\mathbf{y}_k - H\left(M_k\left(\mathbf{x}_0\right)\right)\right] \end{cases}$$

The analysis of 4DVar is converted to a minimization process of the cost function (J) under both the constraints from observations and model trajectory.

Emerging DA techniques: ensemble Kalman filter

EnKF approach propagates and updates uncertainties



Forecast step:

$$\mathbf{x}_i^b = M\mathbf{x}_i^a \ ,$$

$$\mathbf{P}_i^b \approx \frac{1}{K-1}\sum_{k=1}^{m}(\mathbf{x}_k^f - \overline{\mathbf{x}}^f)_i(\mathbf{x}_k^f - \overline{\mathbf{x}}^f)_i^T$$

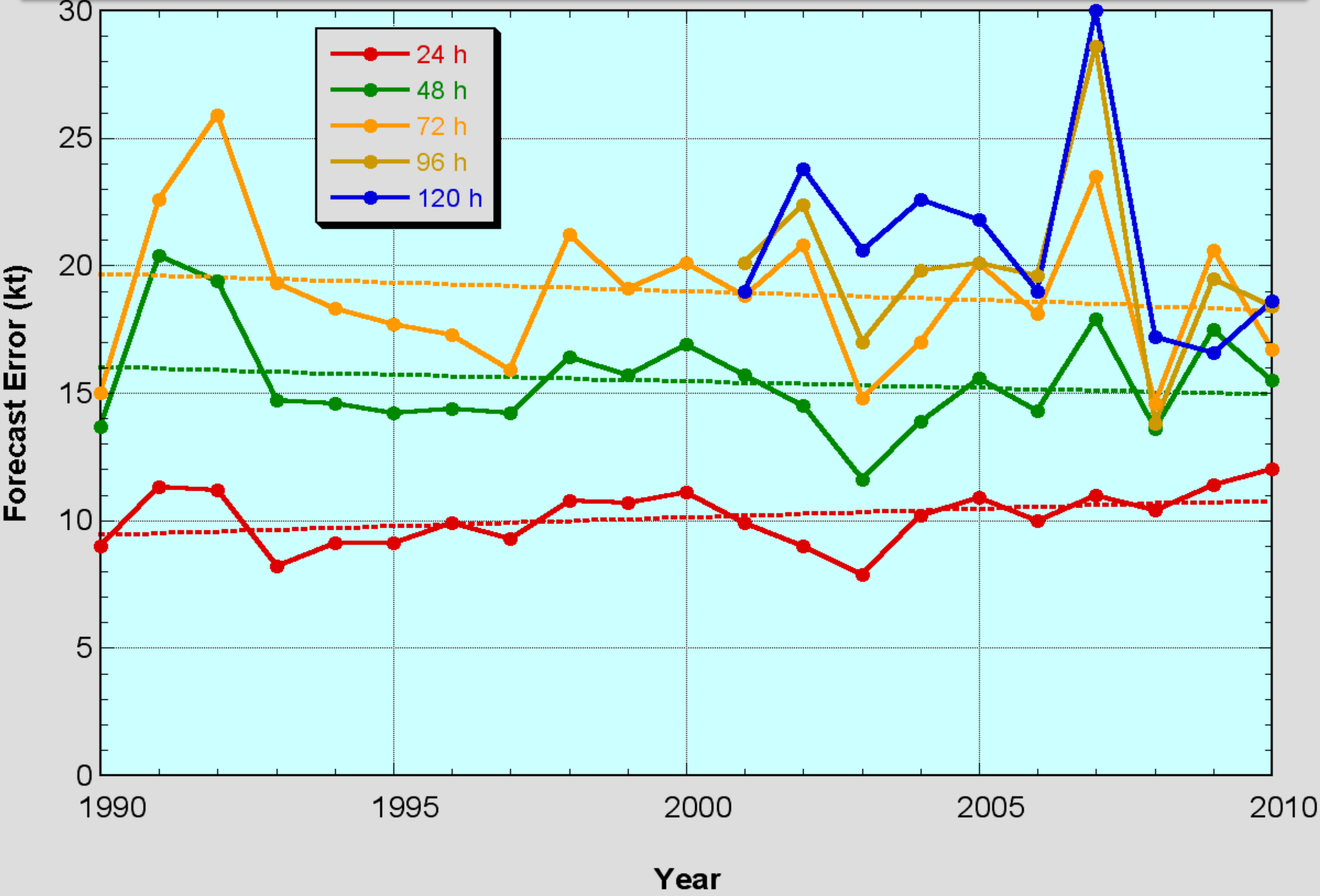
Analysis step:

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K}_i(\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f)$$

$$\mathbf{K}_i = \mathbf{P}_i^b\mathbf{H}^T[\mathbf{H}\mathbf{P}_i^b\mathbf{H}^T + \mathbf{R}]^{-1}$$

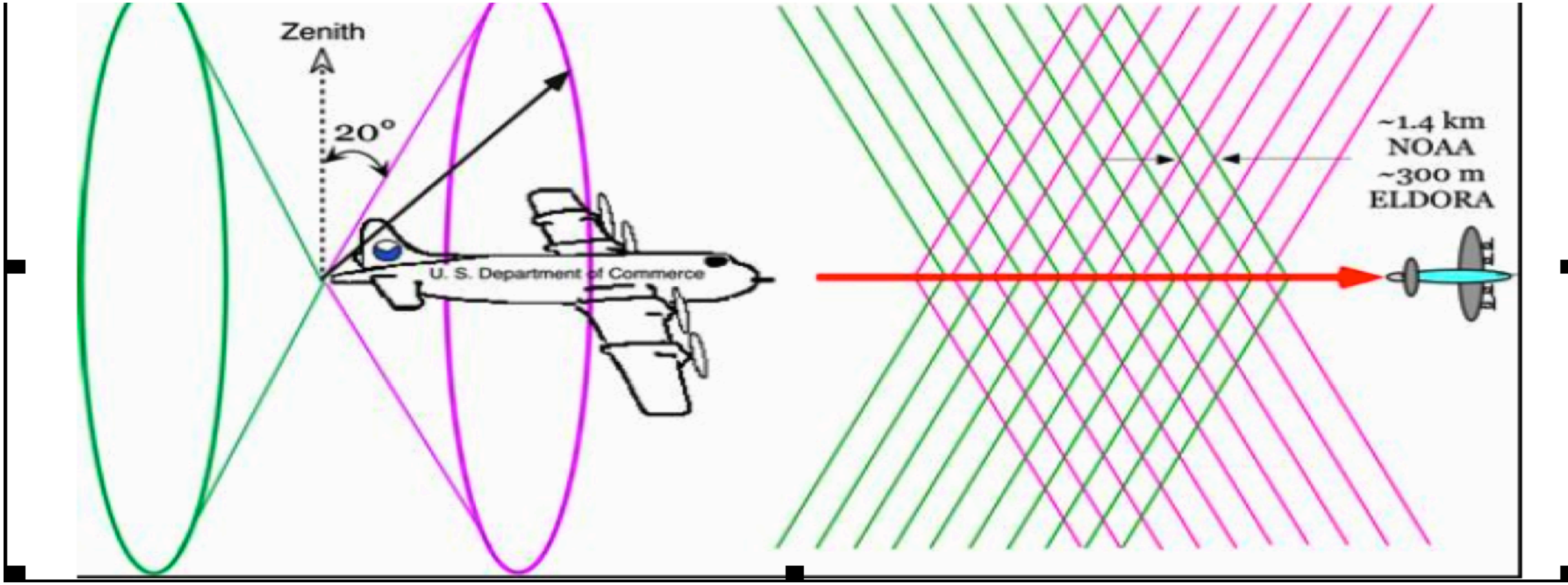
$$\mathbf{P}^a_{i\ n\times n} = [\mathbf{I} - \mathbf{K}_i\mathbf{H}]\mathbf{P}_i^b = [(\mathbf{P}_i^b)^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}$$

National Hurricane Center Official Intensity Errors



Assimilate Airborne Doppler Winds with WRF-EnKF

Available for 20+ years but never used in operational models due to the lack of resolution and/or the lack of efficient data assimilation methods



Superobservations: 1. Separate forward and backward scans; 2. treat every 3 adjacent full scans as one fixed-space radar (translation<5km); 3. thinning ---one bin for 2 km in radial distance and 3 degree in scanning angle; 4. use medium as SO after additional QC checking

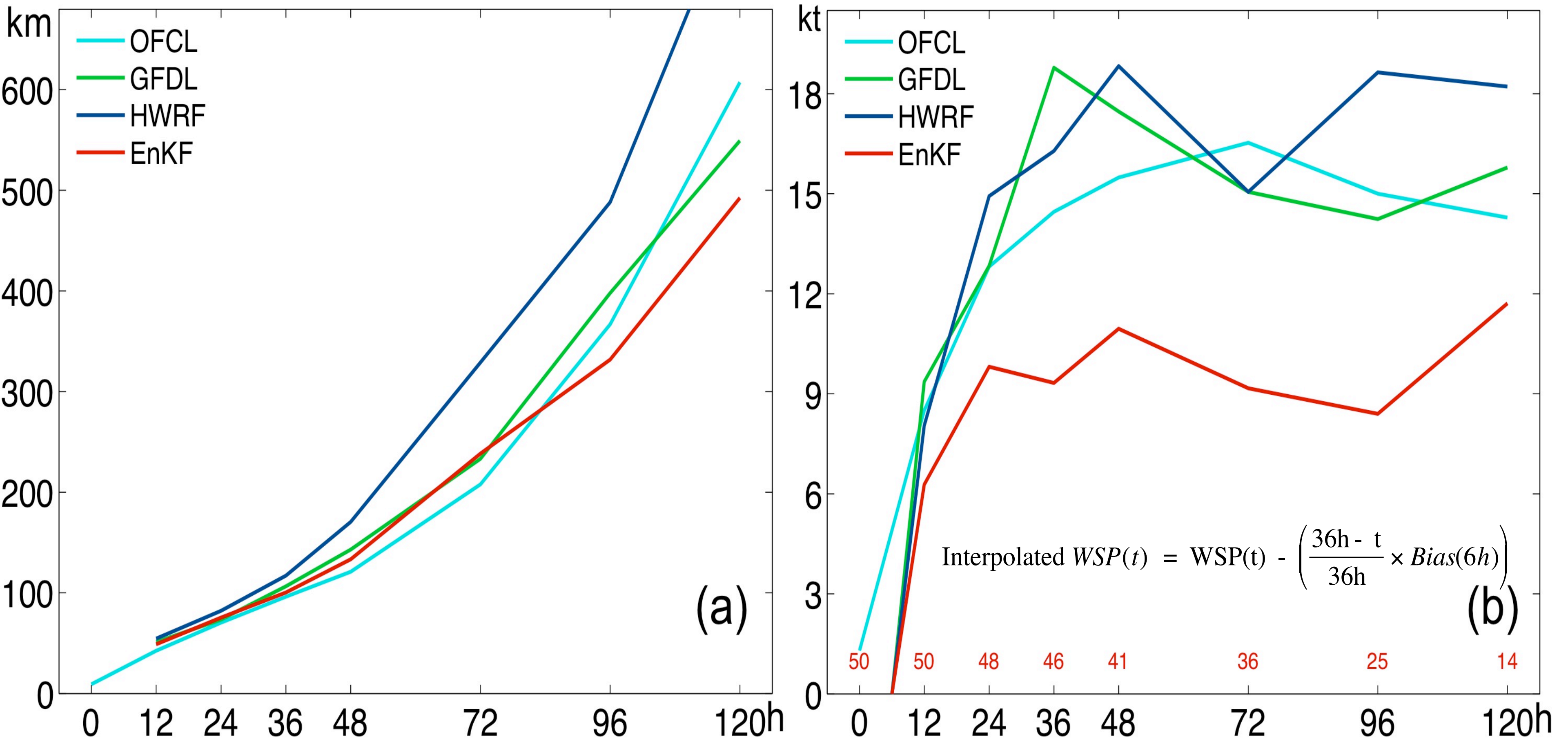
These SOs are generated on flight of NOAA P3's; transmitted to ground in real-time

WRF-EnKF: 3 domains (27, 9, 3km), 60-member ensemble

(Weng and Zhang 2012 MWR)

PSU WRF-EnKF Forecasts Assimilating Airborne Vr

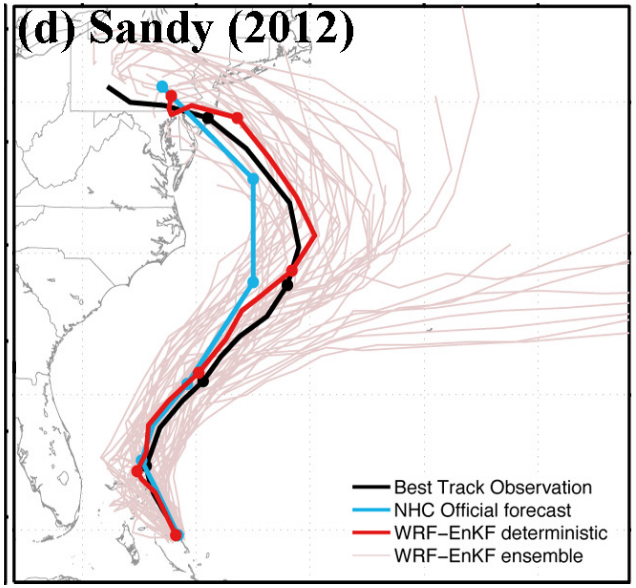
Mean absolute track (km) & intensity (kts) error for all 2008-2010 P3 missions

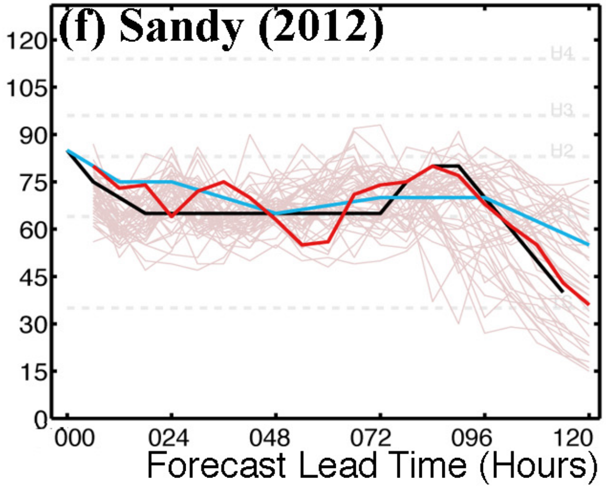


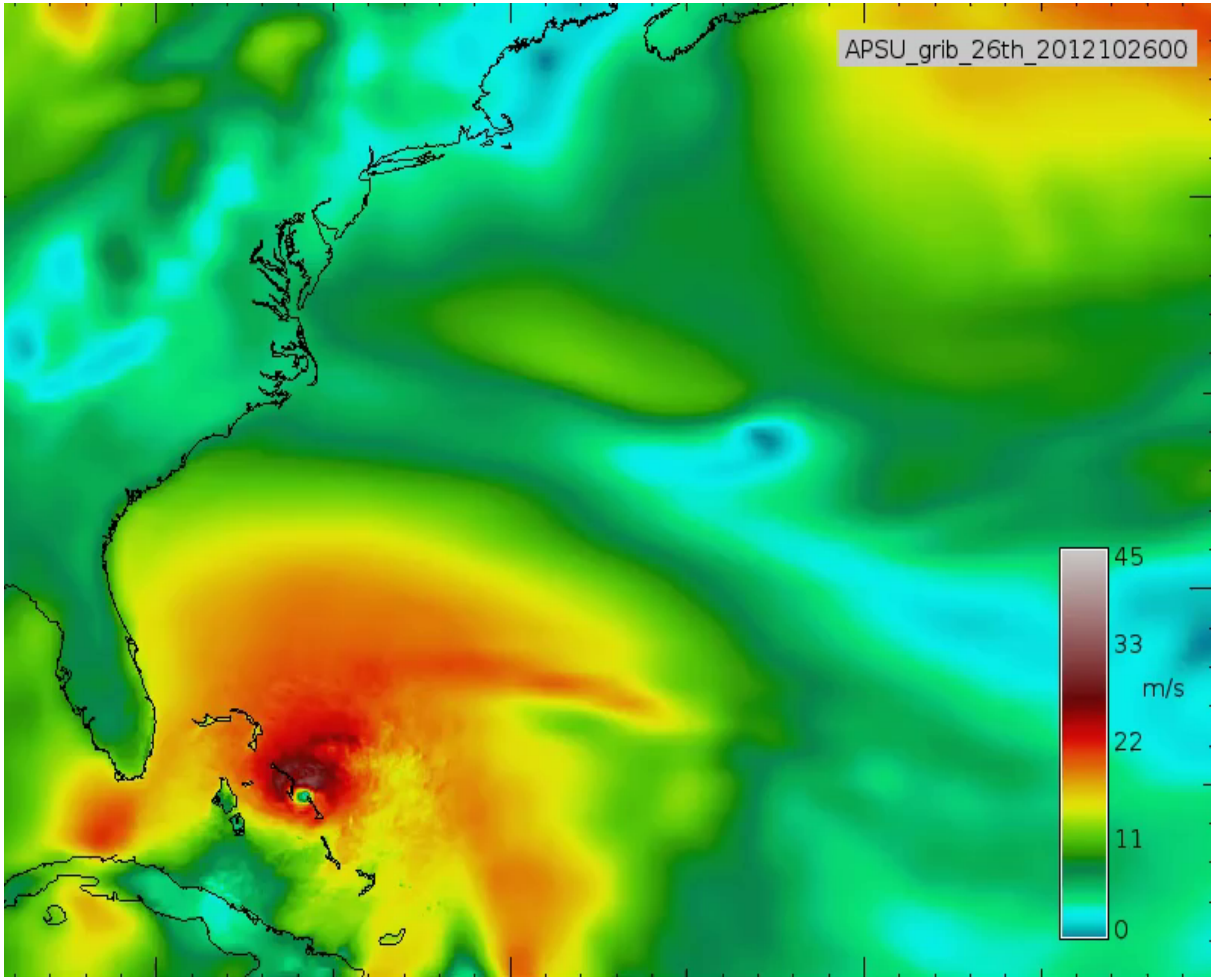
Better resolution & physics; better data (inner core obs); better assimilation (EnKF); big computing

(Zhang et al. 2011 GRL)

PSU Realtime WRF-EnKF w/ assimilation of airborne Doppler winds for Sandy (2012)

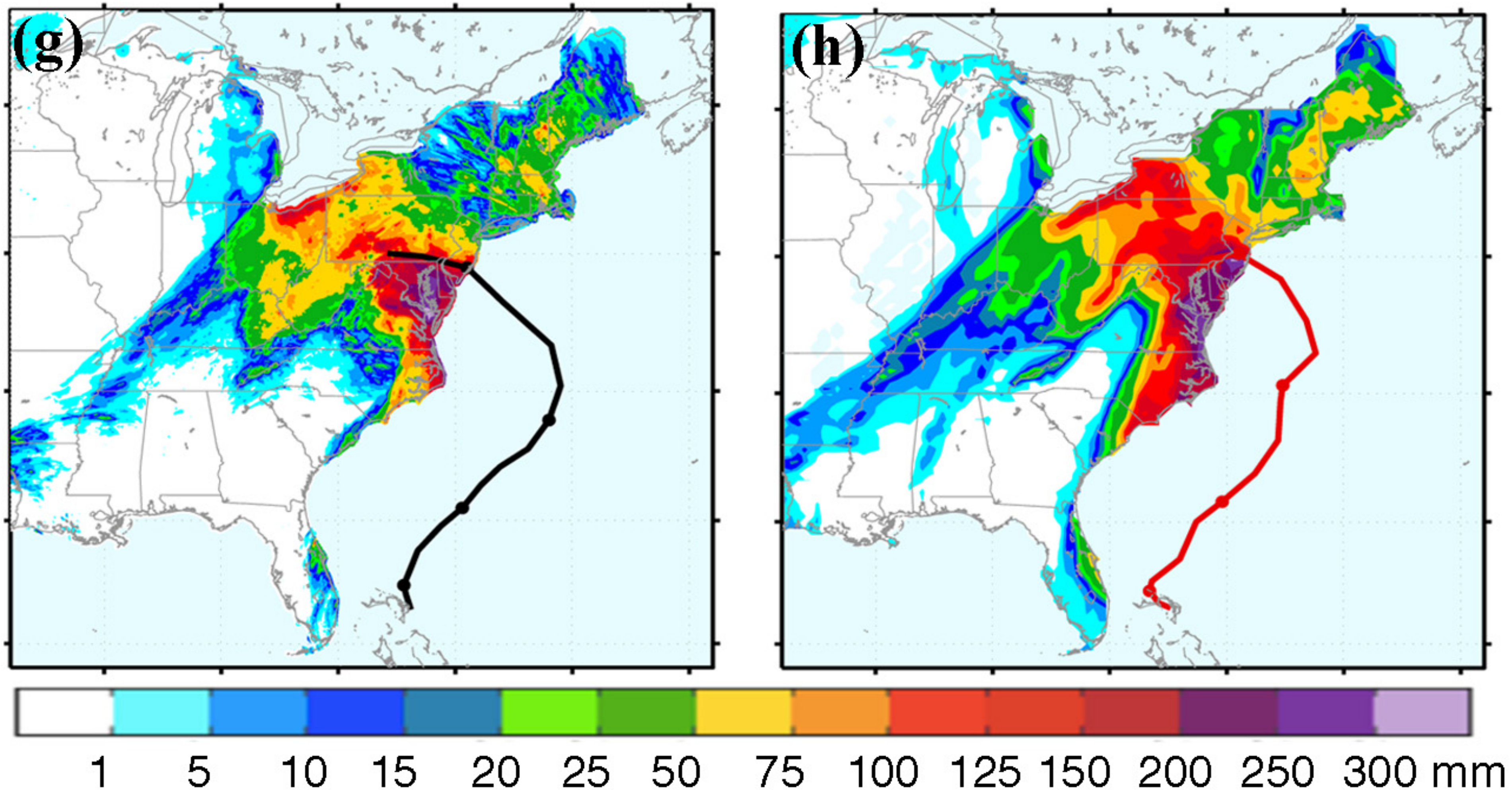






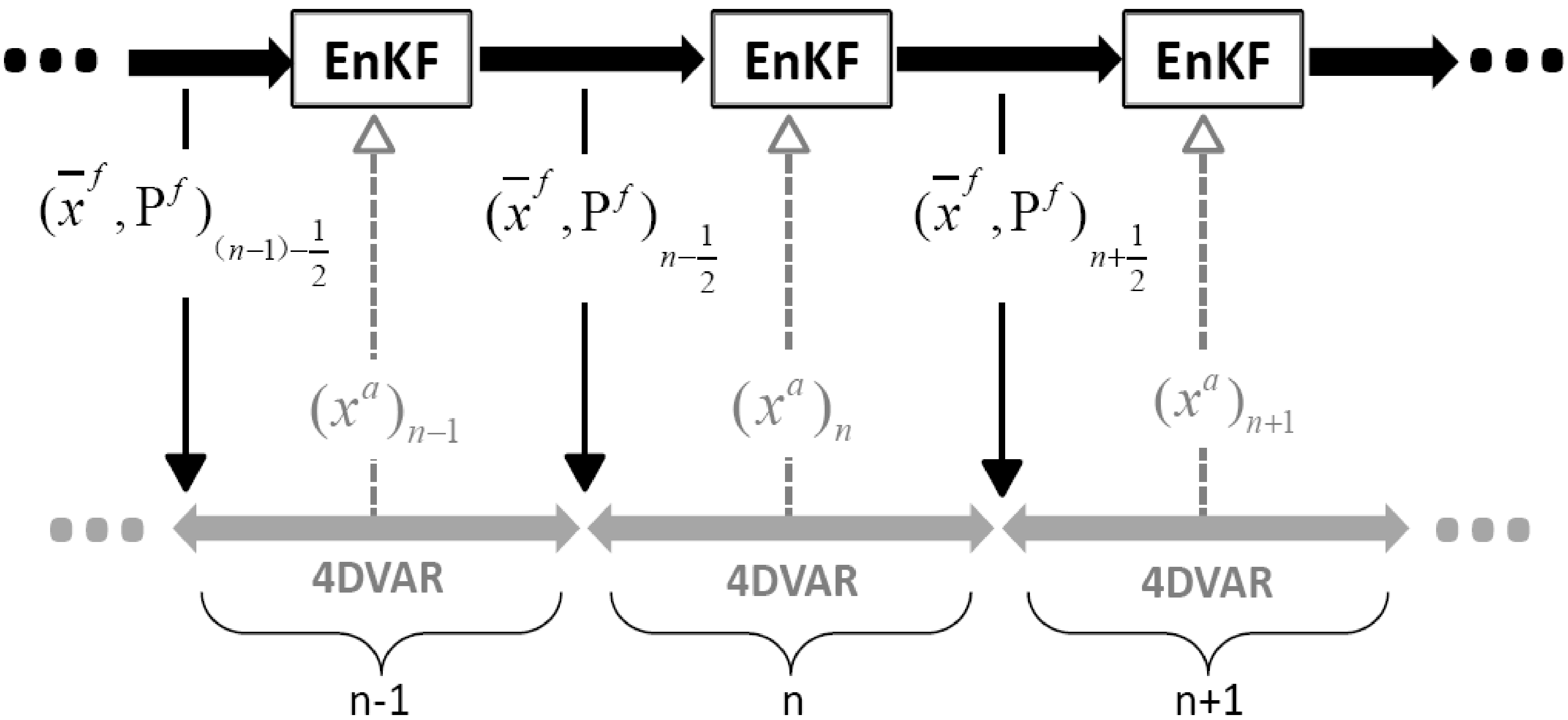
PSU Realtime WRF-EnKF w/ assimilation of airborne Doppler winds for Sandy (2012)

Observation of 5-day total rainfall ***WRF-EnKF forecast***



Emerging DA techniques: hybrid and coupling

E4DVAR: 2-way Coupling of EnKF with 4DVar



Necessary Variable Changes:

EnKF provides ensemble-based background error covariance (*P^f*) for 4DVar

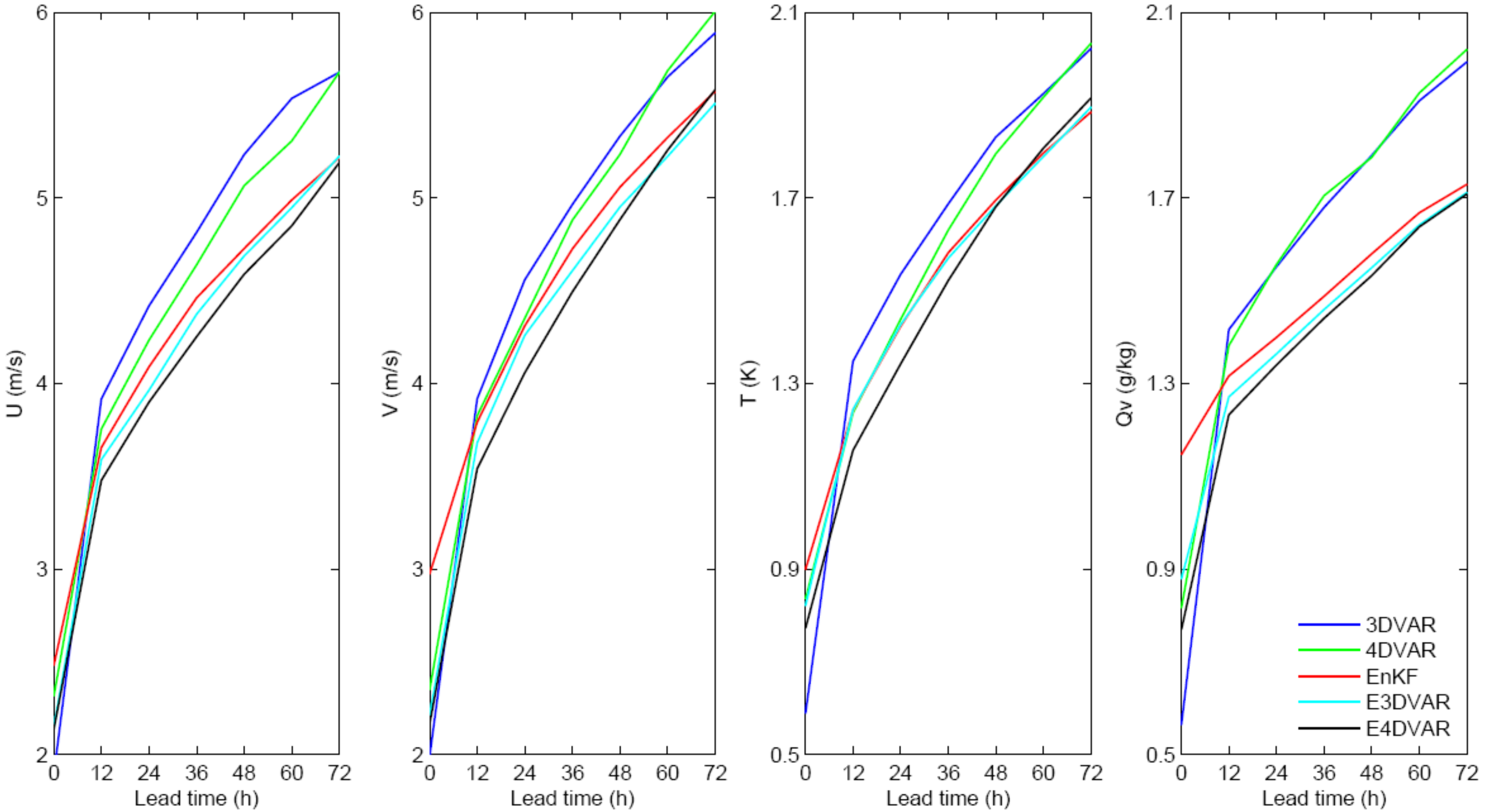
EnKF provides the prior ensemble mean ($\bar{x}^f$) as the first guess for 4DVar

4DVar provides deterministic analysis ($\bar{x}^a$) to replace the posterior ensemble mean for the next ensemble forecast

Proof-of-concept in Zhang, Zhang and Hansen (2009 AAS), real-data expts in Zhang and Zhang (2012 MWR)

E4DVar, E3DVar vs. EnKF, 3DVar, 4DVar

Total RMSE of U, V, T and Q with 0~72 lead time

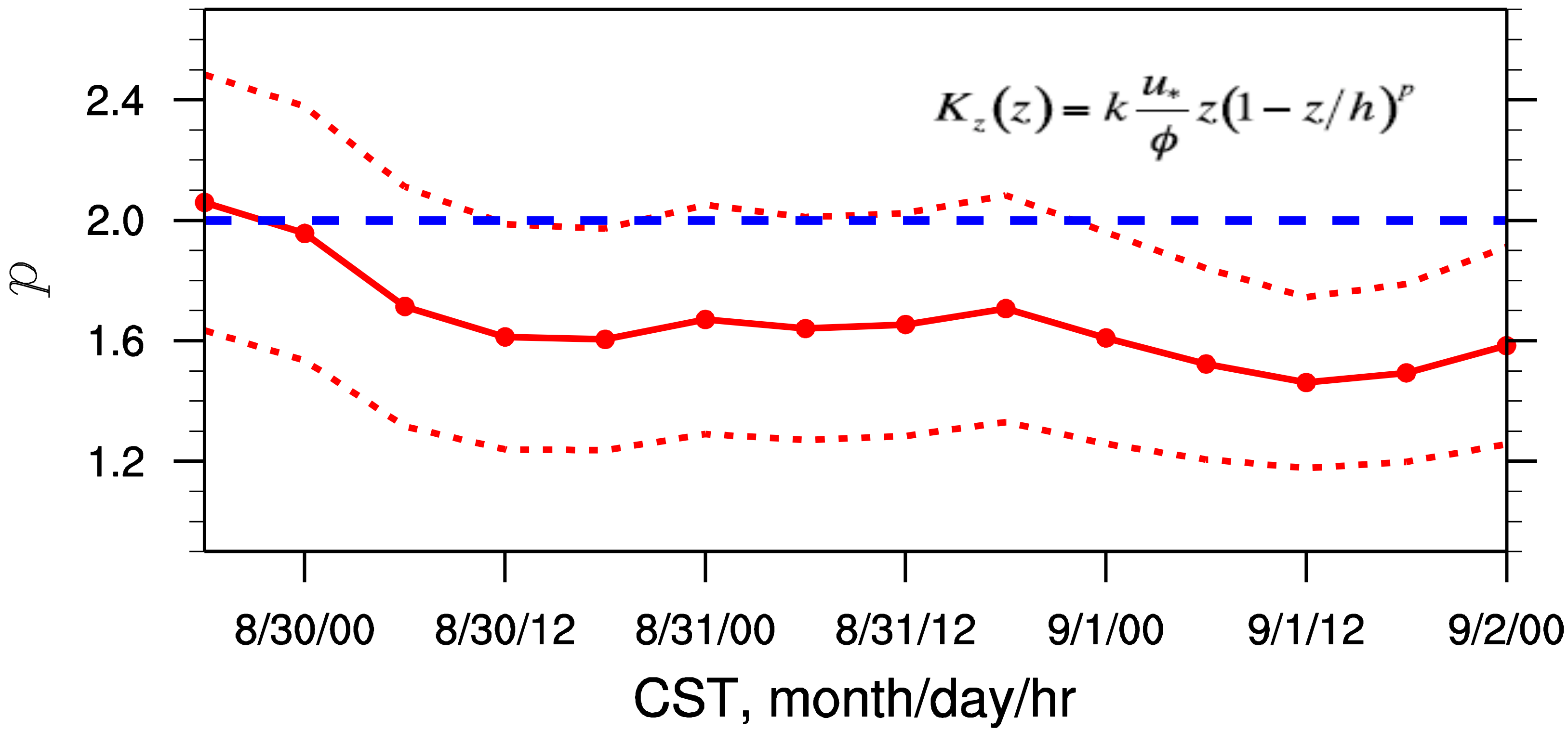


(Zhang and Zhang 2012; Zhang et al. 201

Emerging DA Science: State + Parameter Estimation

Simultaneous State and Parameter Estimation for Treatment of Model Error

A real-data study on WRF/PBL error

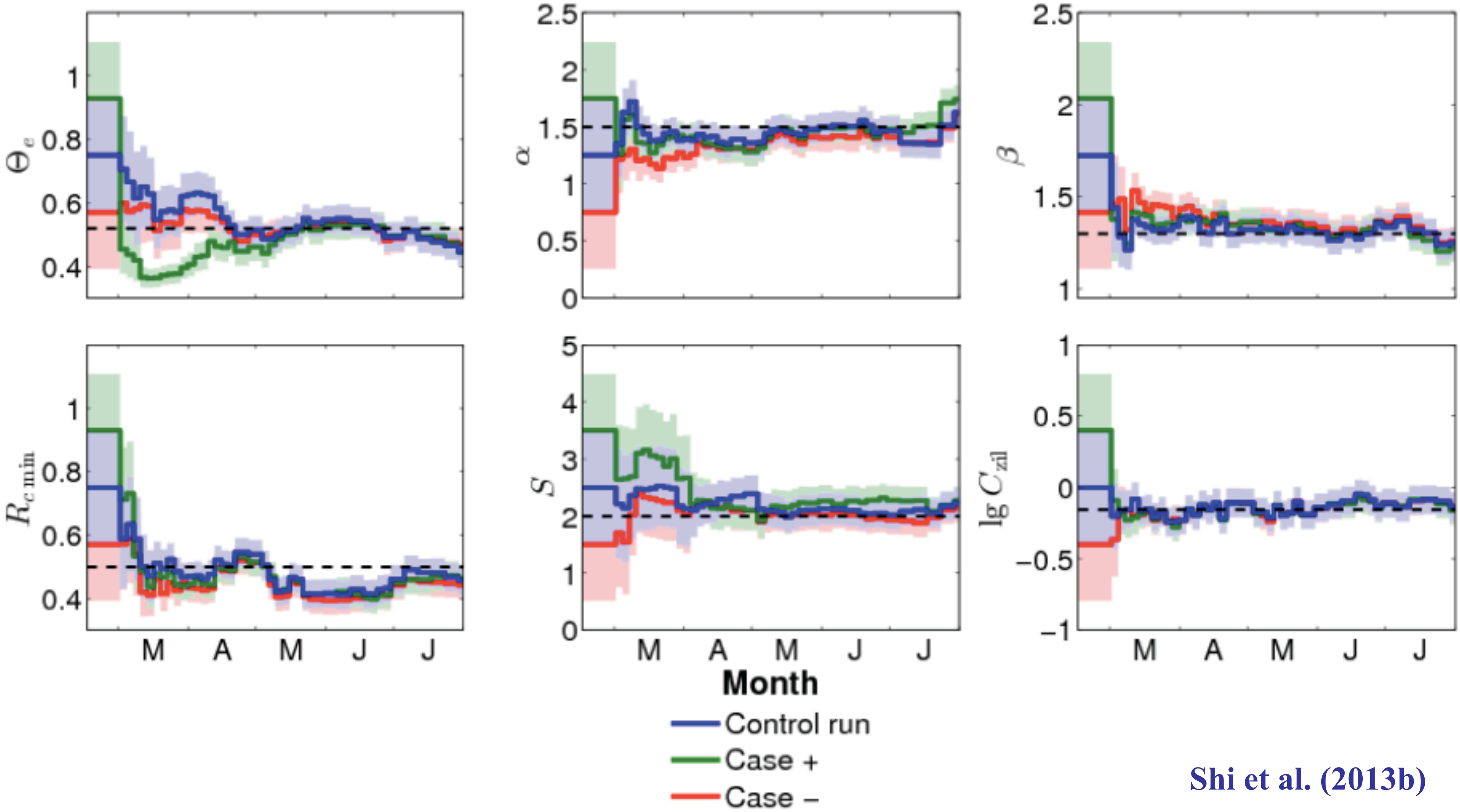


During most of the simulation, SSPE predicts *p* values lower than 2.0 (default). This corresponds to stronger diffusivity in the middle and upper daytime PBL.

Hu, Zhang & Nielsen-Gammon (2010a GRL)

Emerging DA Science: State + Parameter Estimation

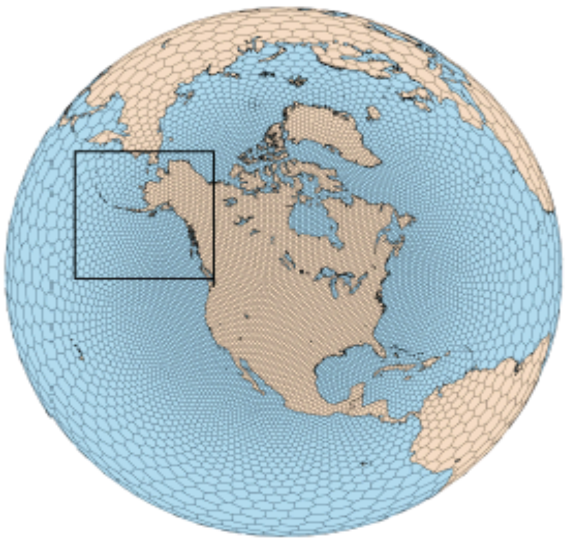
Parameter Estimation of Physically-Based Distributed Land Surface Hydrologic Model Using EnKF

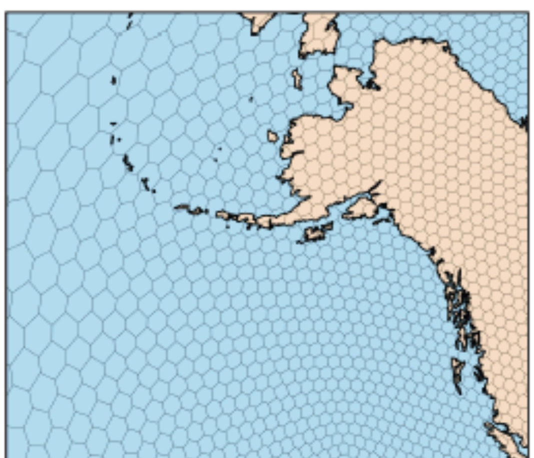


Shi et al. (2013b)

Towards global cloud-resolving models: NCAR/DOE

MPAS
Model for Prediction Across Scales





MPAS-Atmosphere

<u>Unstructured spherical centroidal Voronoi meshes</u>

Mostly *hexagons*, some pentagons and 7-sided cells.
Cell centers are at cell center-of-mass.
Lines connecting cell centers intersect cell edges at right angles.
Lines connecting cell centers are bisected by cell edge.
Mesh generation uses a density function.
Uniform resolution – traditional icosahedral mesh.

<u>C-grid</u>

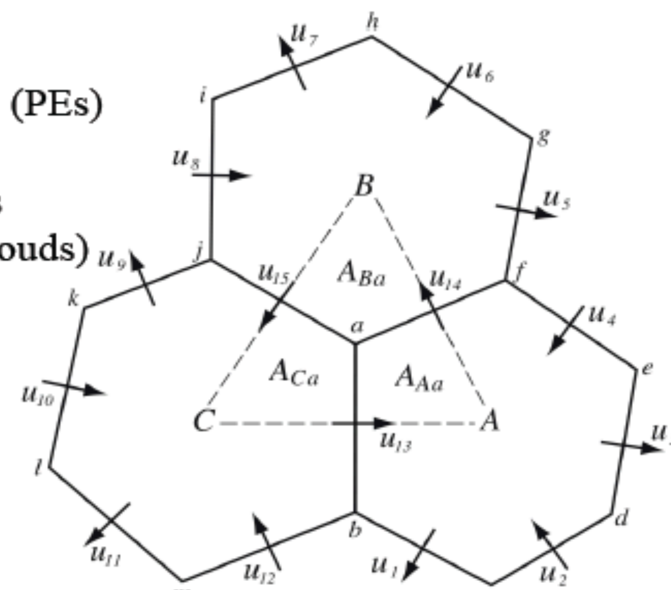
Solve for normal velocities on cell edges.

<u>Solvers</u>

(1) hydrostatic equations (PEs)
(2) Fully compressible nonhydrostatic equations (explicit simulation of clouds)

<u>Solver Technology</u>

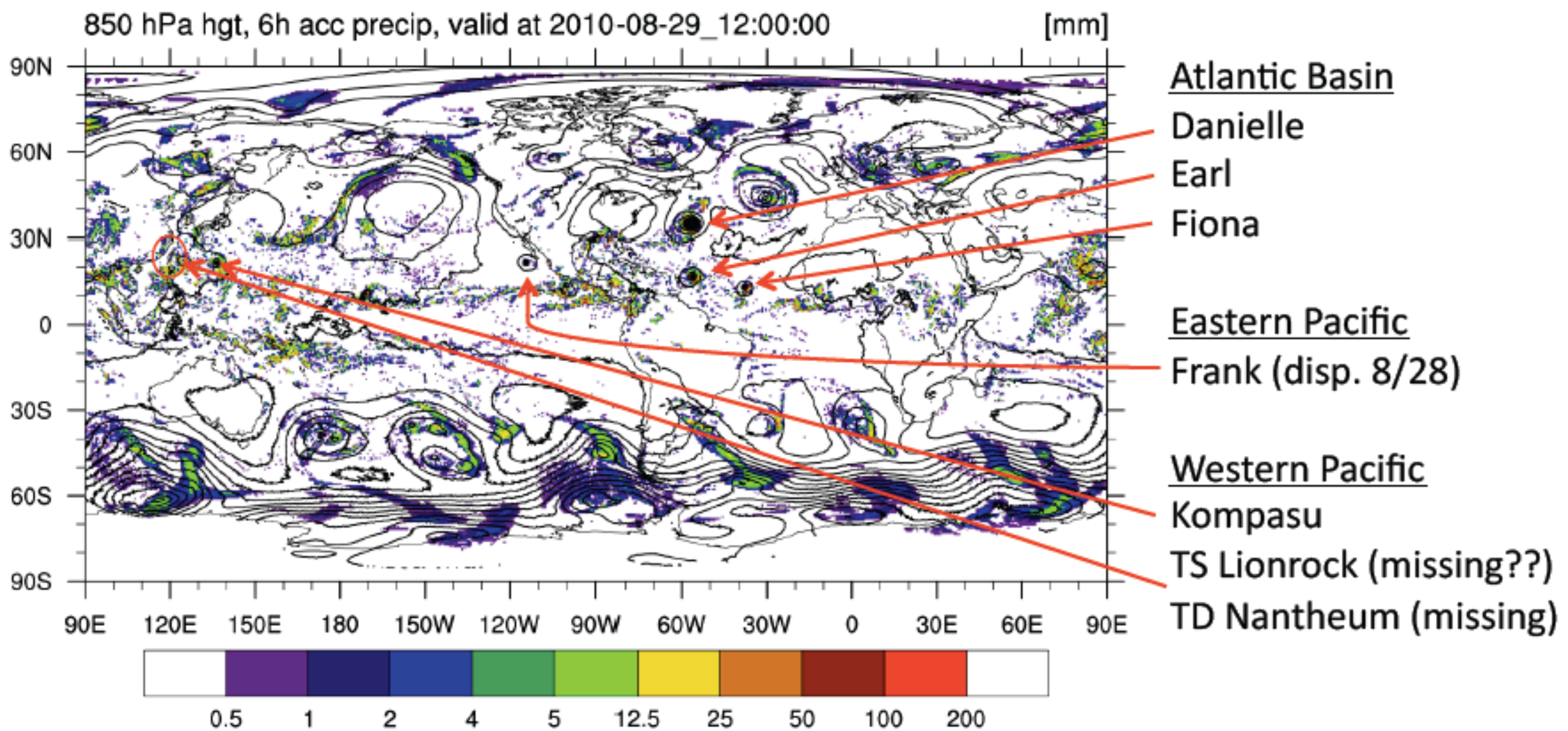
Integration schemes are similar to WRF.



Demands for higher resolution and more computing: Towards cloud-resolving global models

MPAS
Model for Prediction Across Scales

MPAS 3km global simulations, 27 Aug– 1 Sept 2010



EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

Mohan Ramamurthy, Unidata/UCAR
Russ Schumacher, Colorado State University
Fuqing Zhang, Penn State University

Workshop Dates: 17-18 December 2012

Earth Cube Workshop Title: Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction

Important science drivers and challenges

- What are the limits of predictability in the atmosphere? What are the sources of uncertainty/errors, and how do they feed into predictability?
- What observations are critically needed to enhance atmospheric predictions, and where? What is the optimal configuration of the observation network?
- What are the appropriate types, combinations, and configurations of parameterization schemes for high-resolution global cloud-resolving models? How can the errors and biases in these parameterizations be quantified and corrected?
- What is the optimal ensemble configuration to accurately predict the distribution of possible outcomes? How many ensemble members are needed and how should the ensembles be initialized?
- What are the advantages and disadvantage of variational versus ensemble-based data assimilation techniques, as well as different types of hybrid and coupling data assimilation approaches?
- What are the most effective ways to post-process ensemble forecasts to achieve reliable and calibrated probabilistic predictions?

Challenges to high-impact, interdisciplinary science

- Significant barriers exist in using the data efficiently or integrating them into data assimilation or ensemble prediction systems. Too much overhead to doing research efficiently – e.g., setting up one's data and analyzing it.
- The scientific community lacks easy-to-use common cyberinfrastructure frameworks, data format standards, sufficient metadata for observations, and methods/tools for quality controlling observations, mining of large volumes of data, visualization, and verification.
- While many good shared facilities exist in this field, their operations and services are not always well coordinated or integrated.
- Lack of a central repository for finding, accessing, and using data and software.
- Significant spin-up time for students in preparing, using, processing, and analyzing data. While similar challenges exist for researchers, such problems are particularly acute for students who have a limited time before they graduate.
- Barriers to collaboration between closely linked disciplines; e.g., Atmospheric Sciences, Computer Science, Mathematics and Statistics

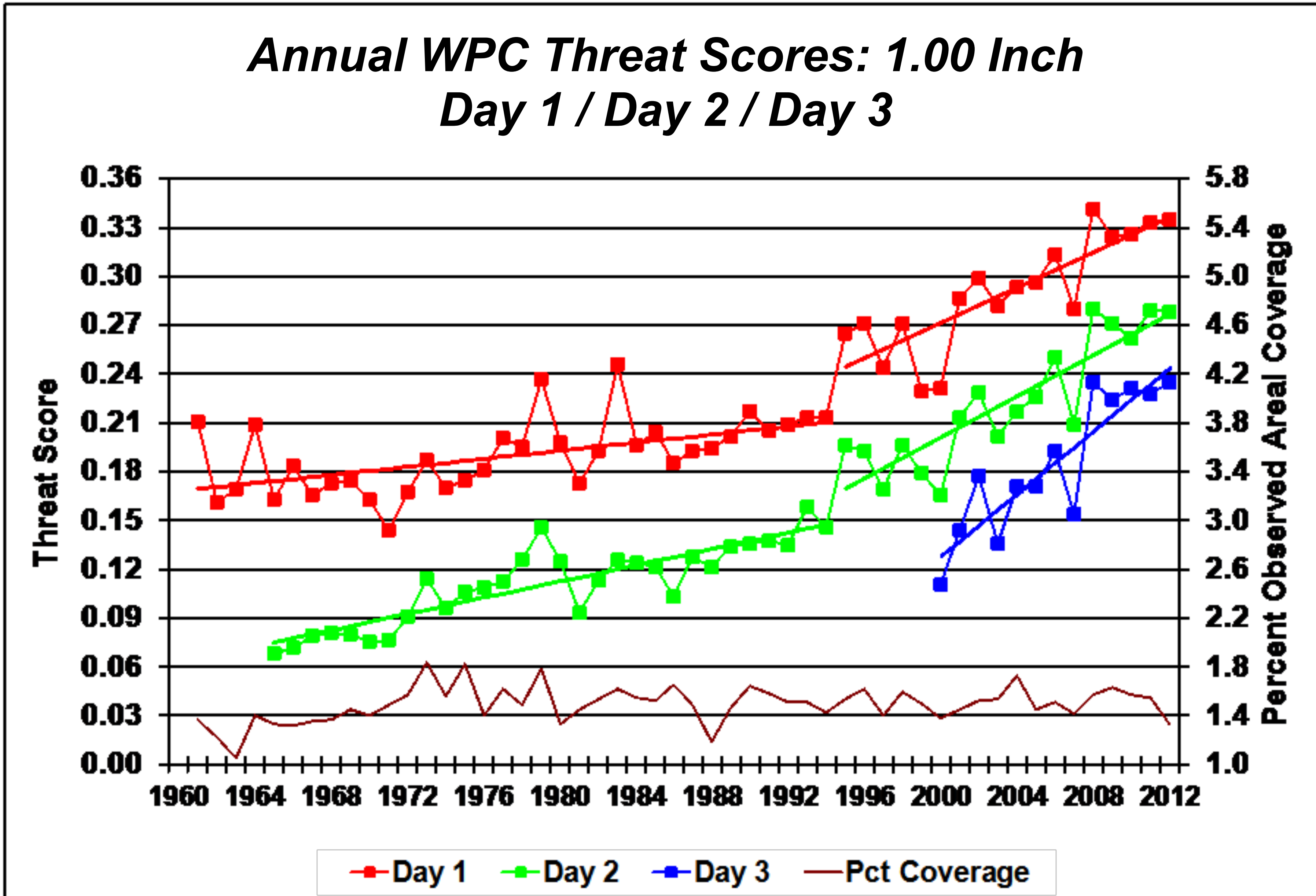
Needed and desired tools, databases, etc.

- Centralized data repositories and services that link existing and future data systems. For example, a centralized community repository could be created for data submission and sharing.
- Advanced software, tools kits, and services for quality control, in-depth data analysis, visualization, verification, and mining of data (observational and model output). These tools and services need to be user-friendly and accessible by the whole scientific community.
- Common data formats and frameworks for assimilation, modeling, analysis and visualization.
- Common data assimilation framework; currently, each assimilation system uses its own framework for data I/O, processing, and running algorithms.
- Collaboration tools, platforms, and frameworks (e.g., Wiki for data)
- Server-side processing tools for data processing, analysis, visualization

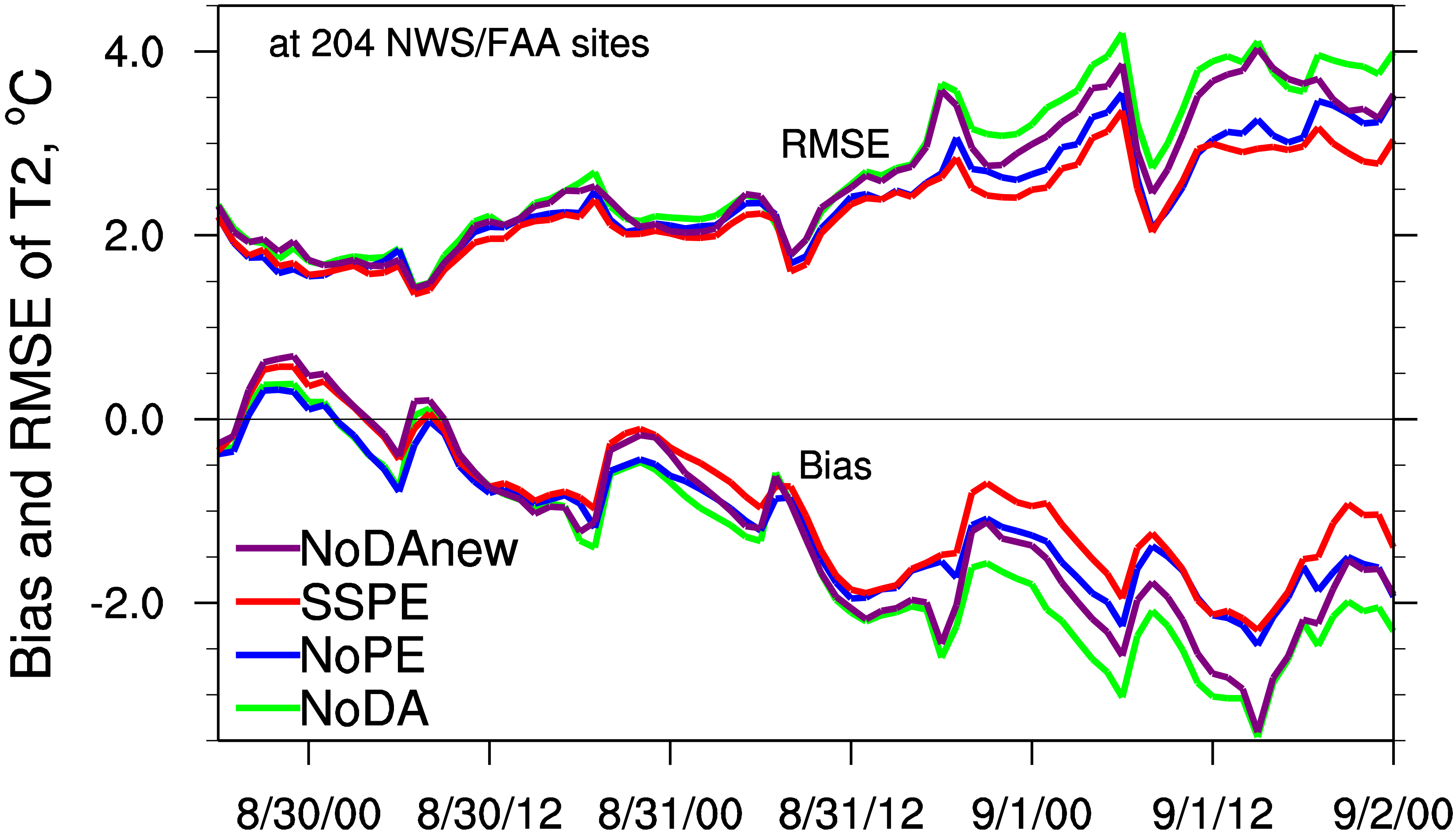
How to move forward using EarthCube

- A pilot project on coordinated, distributed national ensemble prediction that involves universities that are interested in participating
- Developing a prototype system that links data sets/systems together, such as the reanalysis data sets; develop a system that works seamlessly, and then expand to include other data sets/systems
- Continued discussion with the goal of developing a concrete plan for greater coordination of ongoing and future programs and facilities that serve the data assimilation and prediction communities, and developing a next-generation testbed facility to advance the science.
- PI meetings to leverage and expand communication, and enhance data sharing, and facilitate sustained interactions
- Entrain current undergraduate and graduate students into research and educational activities related to "big data", ensemble prediction and data assimilation, and EarthCube, to move these initiatives forward for the future scientific workforce.
- Reach out to other geoscience communities, including oceanography, and hydrology, as well as the computer and information science communities.

How Far We've Come



Bias and error of T2

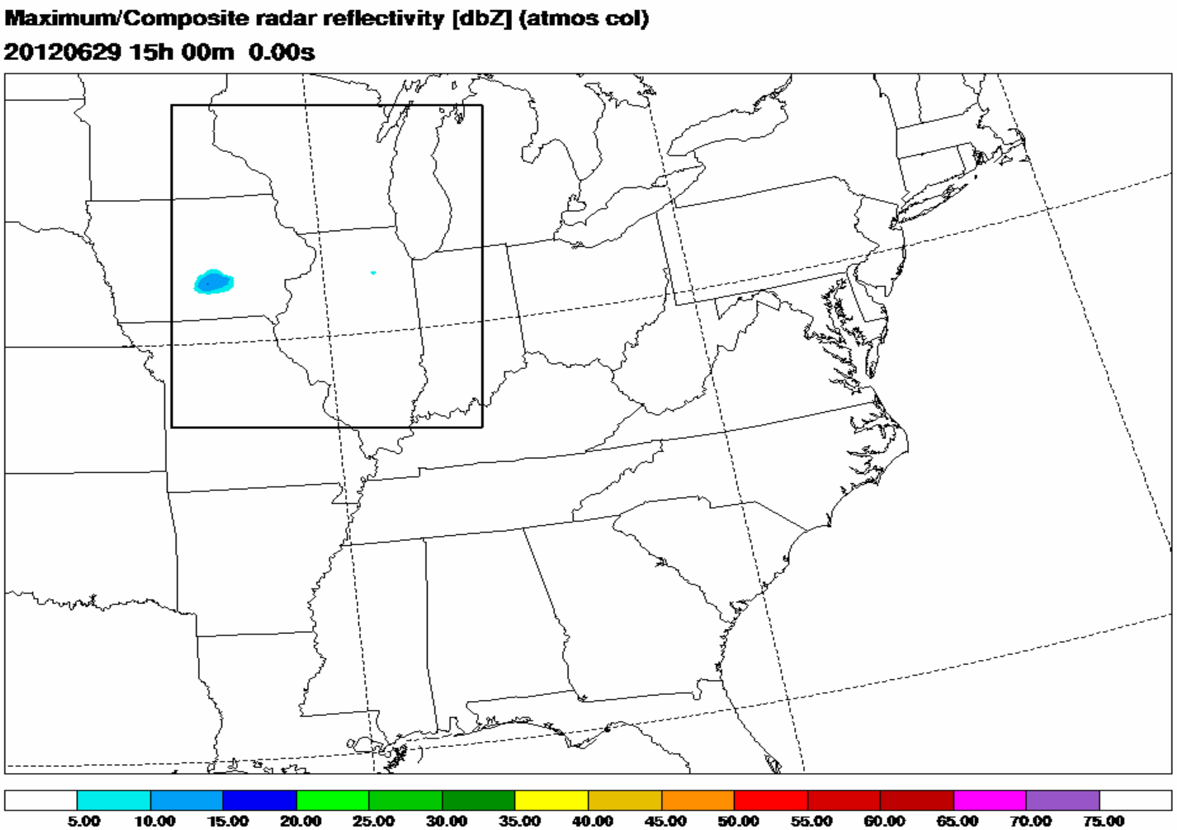


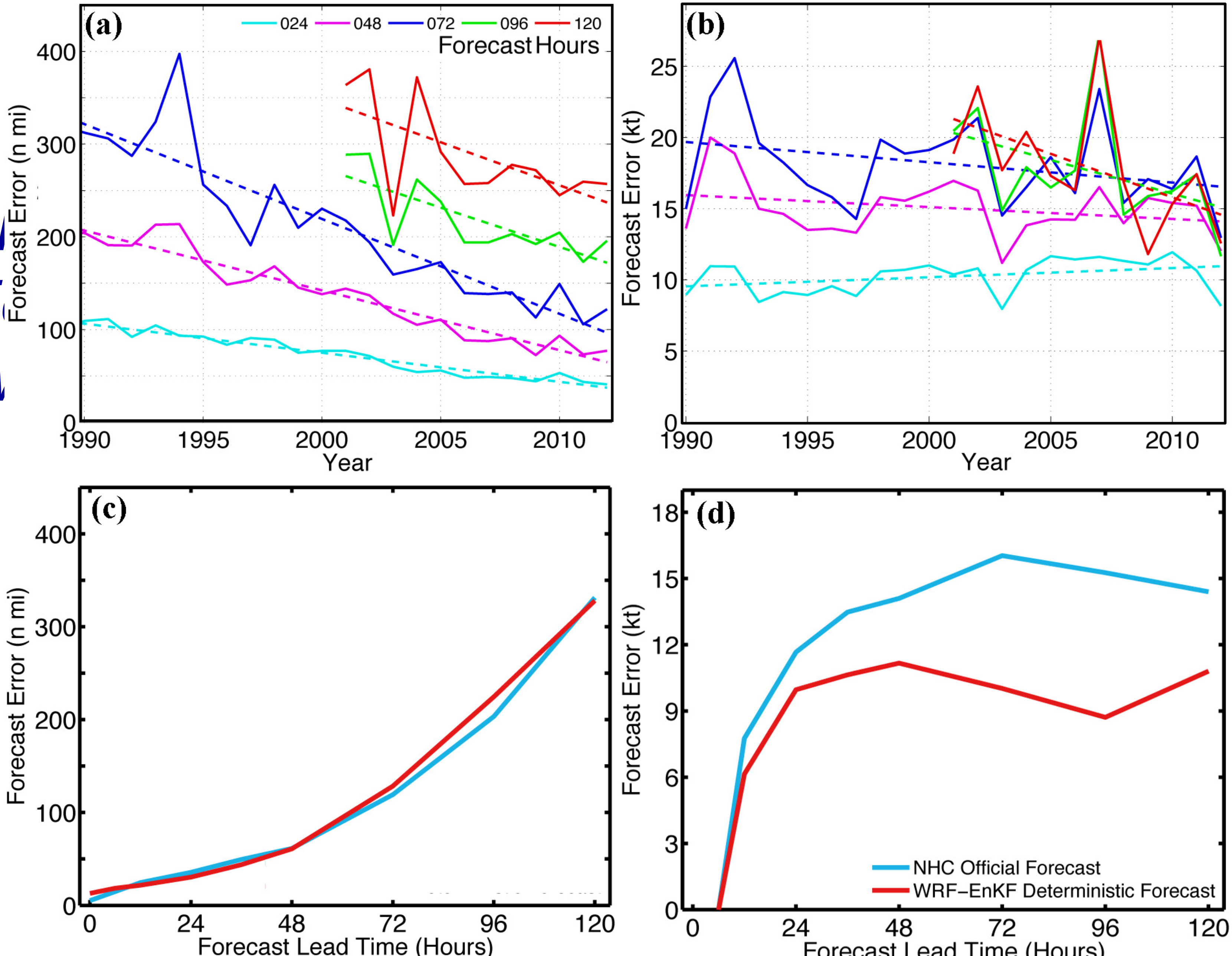
Hu, Zhang & Nielsen-Gammon (2010a GRL)

NMMB storm-scale nest: Application to High Impact Weather

1.3 KM NMMB Moving Nest applied to 29 June 2012 DC Derecho
Simulated maximum composite radar reflectivity

- **Initialized with RAPv2 analysis**
- **12km parent domain with 1.3km moving nest**
- **Nest motion prescribed**
- **Output every 10 minutes**
- **Supports NOAA WoF and WRN inititaves**
- **Ongoing development:**
  - Nest movement based on phenomena of interest
  - 2-way nesting for hurricane applications
  - Computationally efficient





(a)
024
048
072
096
120
Forecast Hours
Forecast Error (n mi)
Year
(b)
Forecast Error (kt)
Year
(c)
Forecast Error (n mi)
Forecast Lead Time (Hours)
(d)
Forecast Error (kt)
Forecast Lead Time (Hours)
NHC Official Forecast
WRF–EnKF Deterministic Forecast

Demands for higher resolution and more computing: results from one idealized large-eddy simulation

Idealized TC:

f-plane

zero env wind

fixed SST

Nested Grids

WRF Model Physics:

WSM3 simple ice

No radiation

Relax to initial temp.

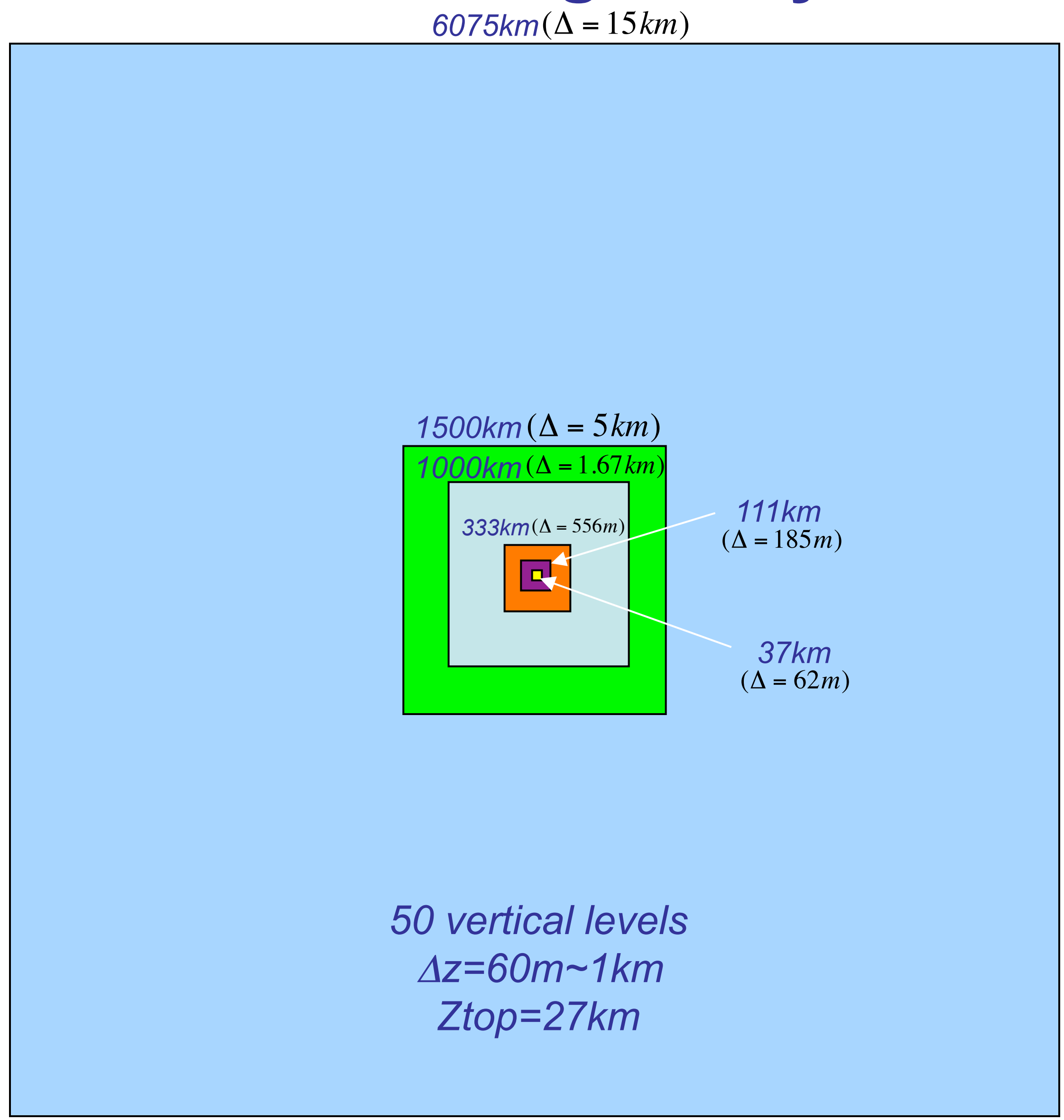
Cd (Donelan)

Ce (Carlson-Boland)

Ce/Cd ~ 0.65

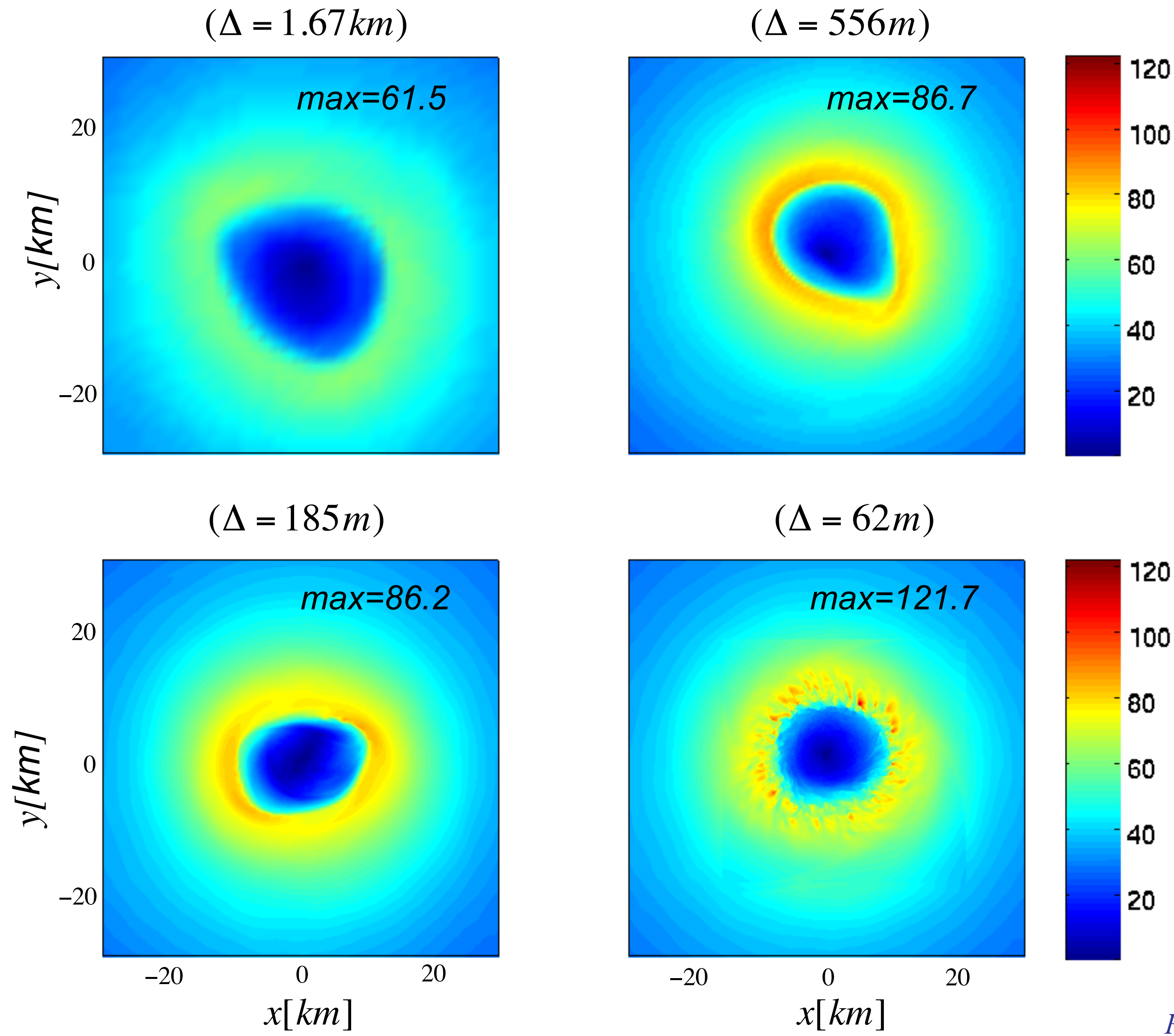
YSU PBL $(\Delta \geq 1.67km)$

LES PBL $(\Delta < 1.67km)$



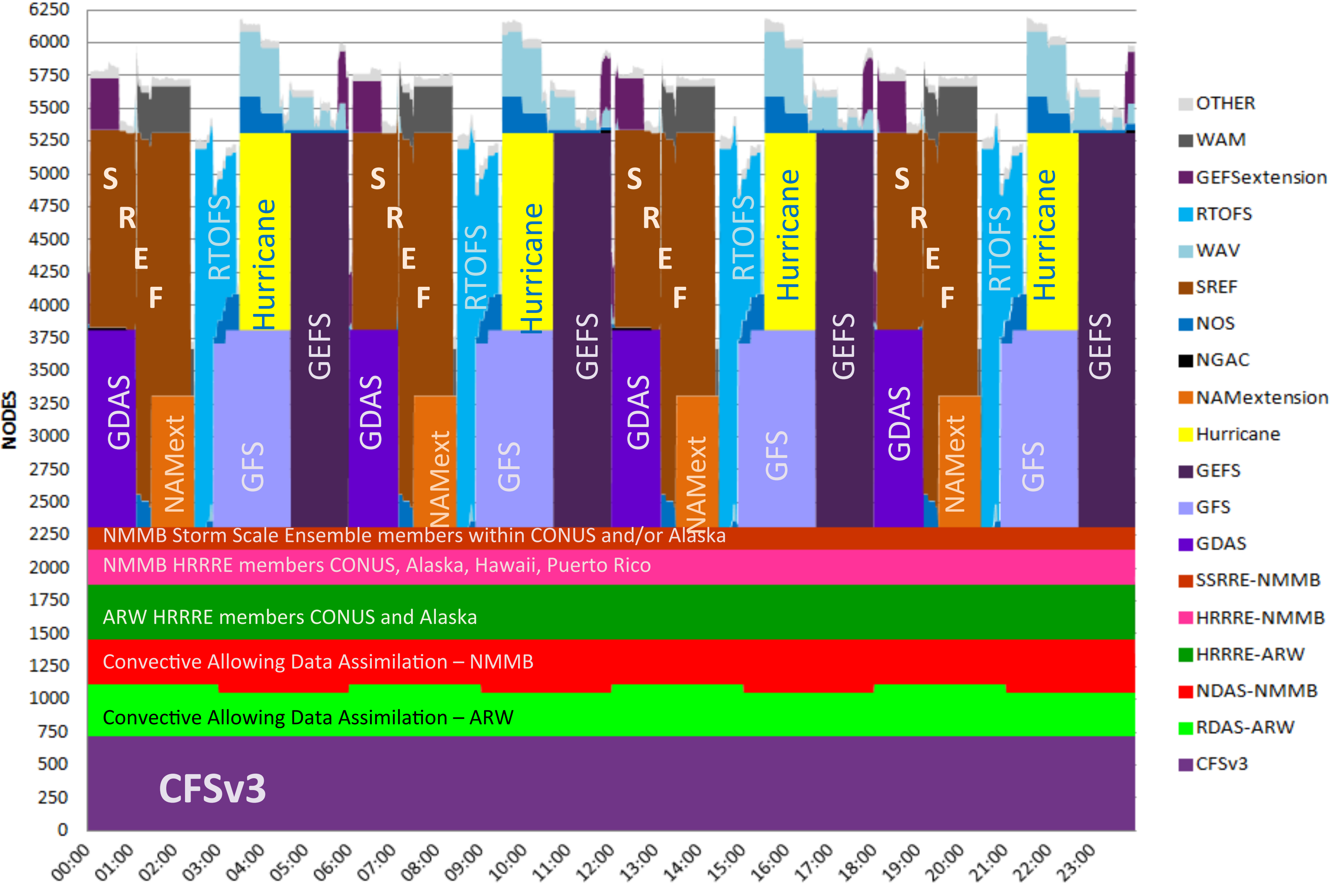
Courtesy of Rich Rotunno at NCAR

Sensitivity to grid resolution: 10-m Wind Speed t=9.75d

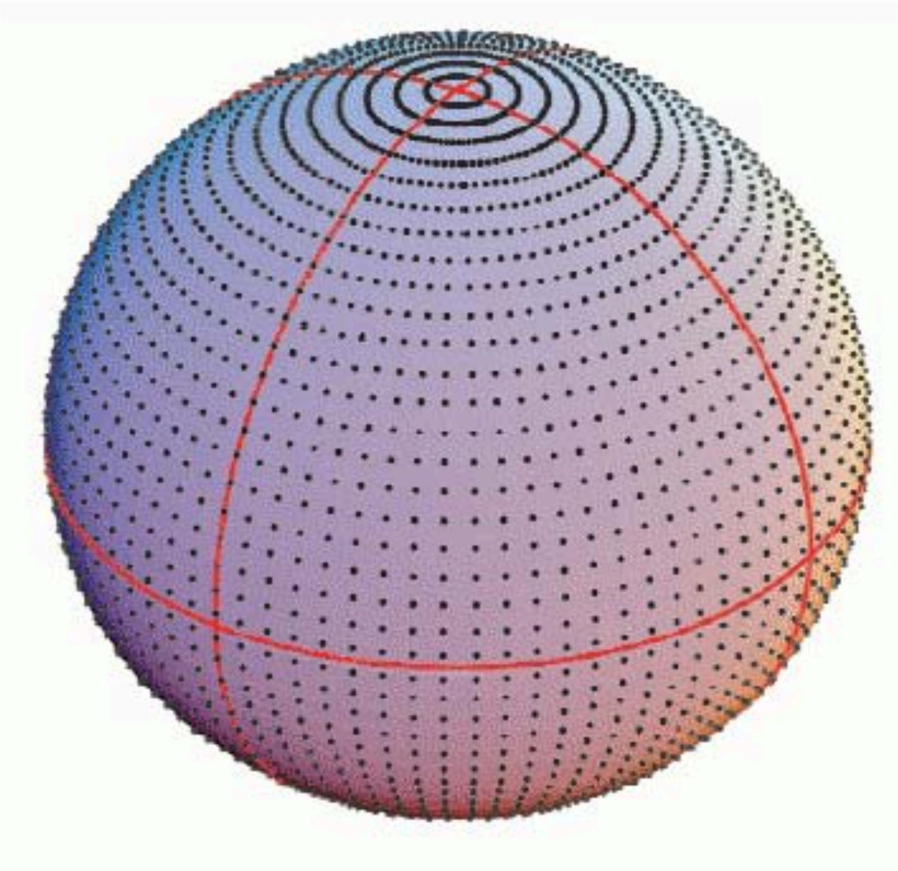


Courtesy of Rich Rotunno at NCAR

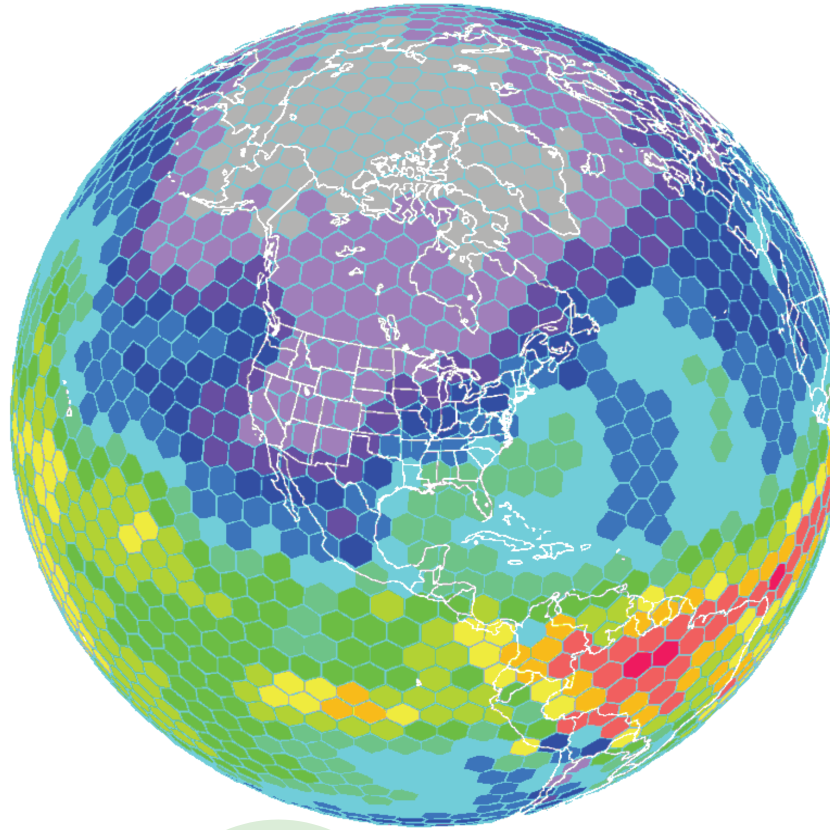
Projected WCOSS Phase 2 (2 Petaflop) End State 2018



Lat/Lon Model

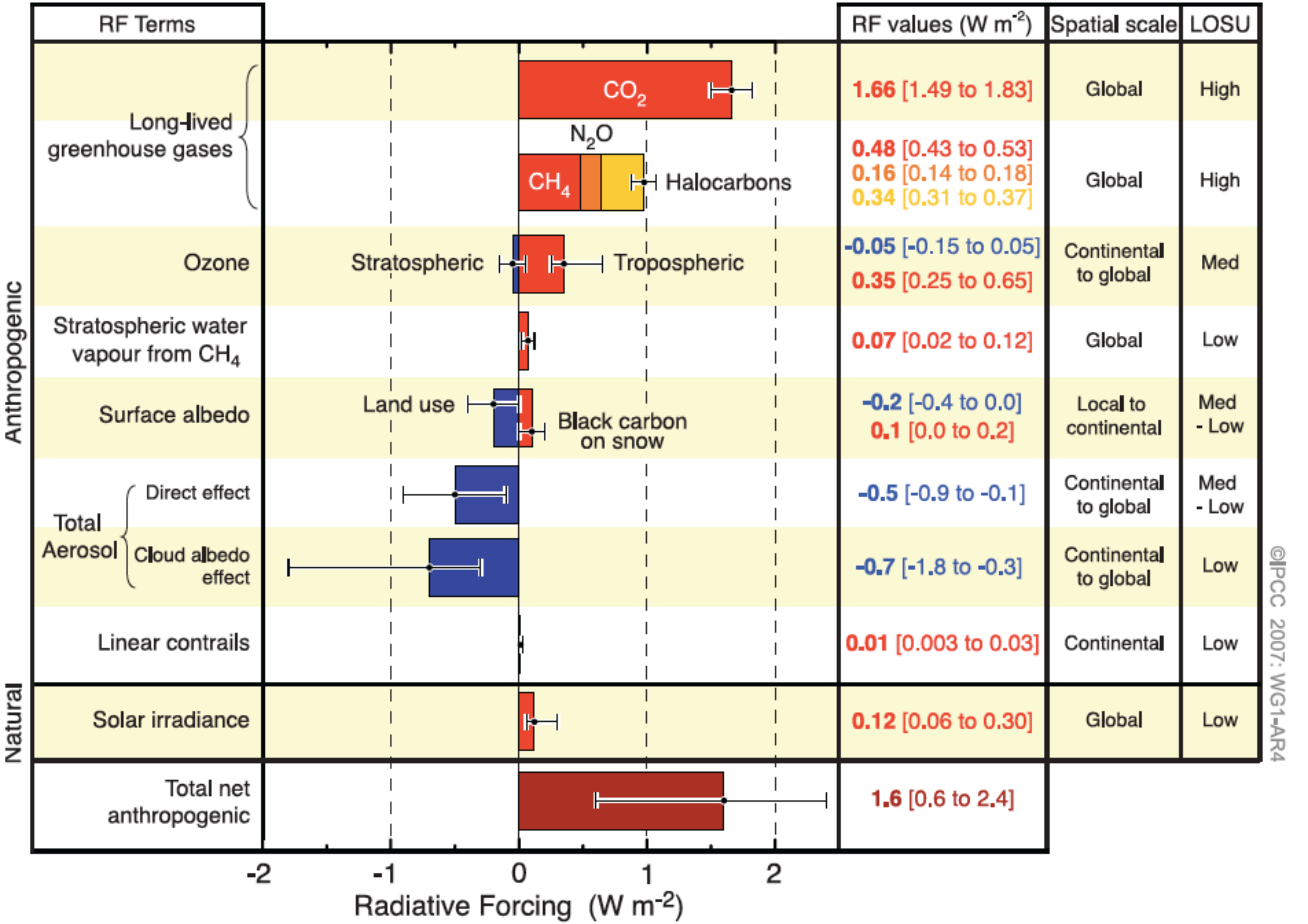


Icosahedral Model

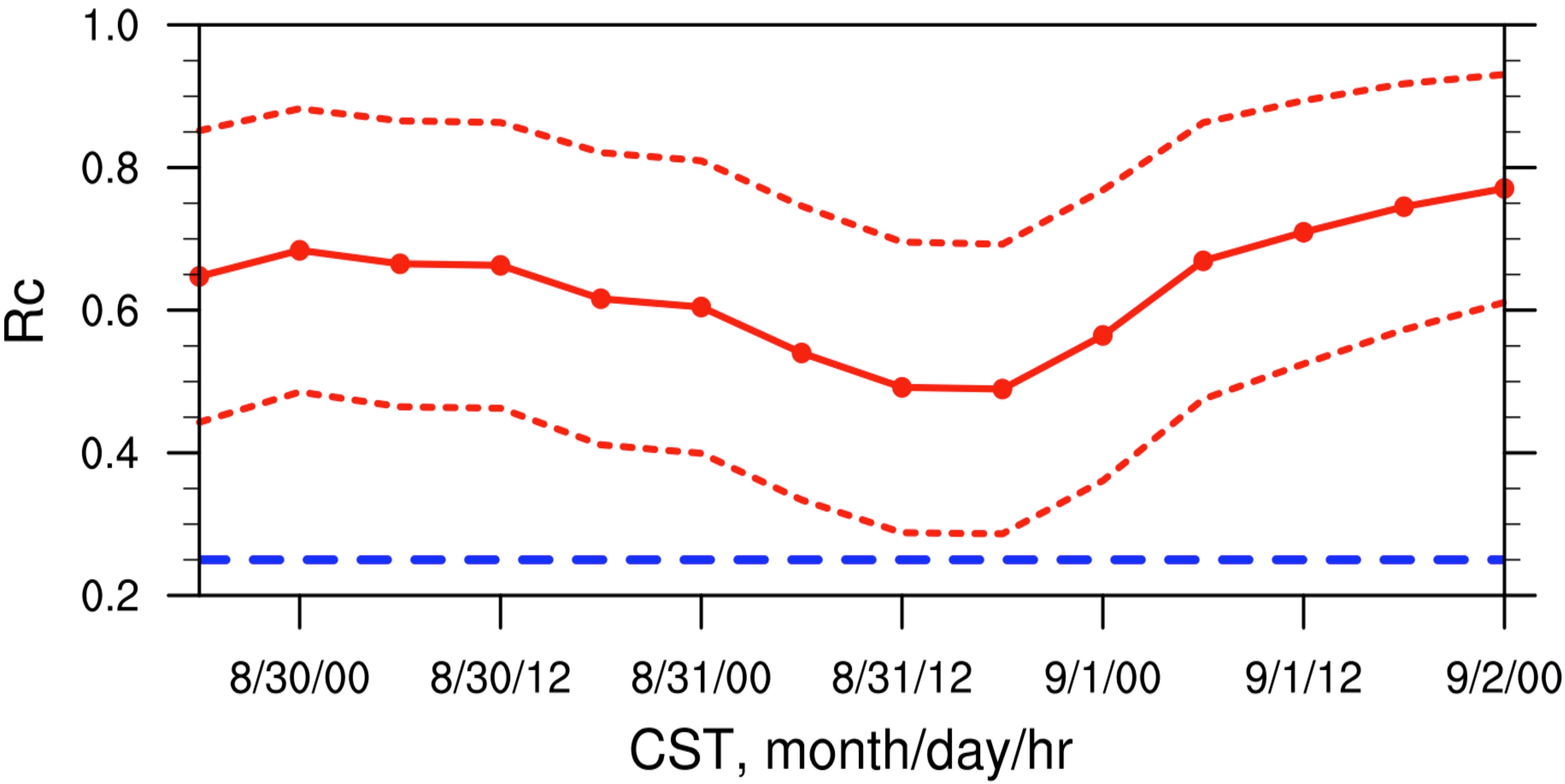


- *Near constant resolution over the globe*
- *Efficient high resolution simulations*

RADIATIVE FORCING COMPONENTS



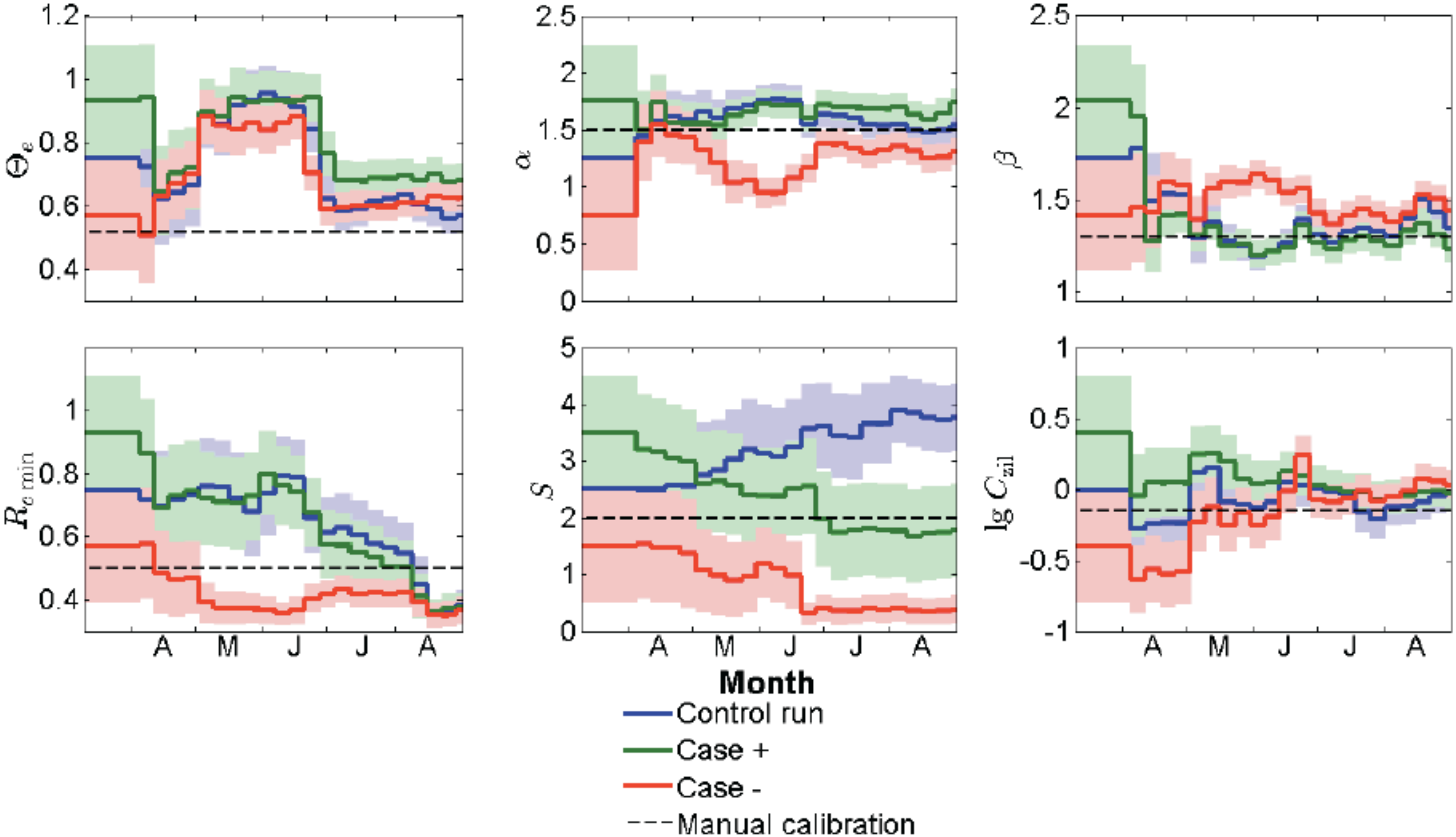
Evolution of *Rc*



During the entire simulation, SSPE predicts *Rc* values higher than 0.25 (default). This corresponds to stronger mixing under weakly stable conditions.

Hu, Zhang & Nielsen-Gammon (2010a GRL)

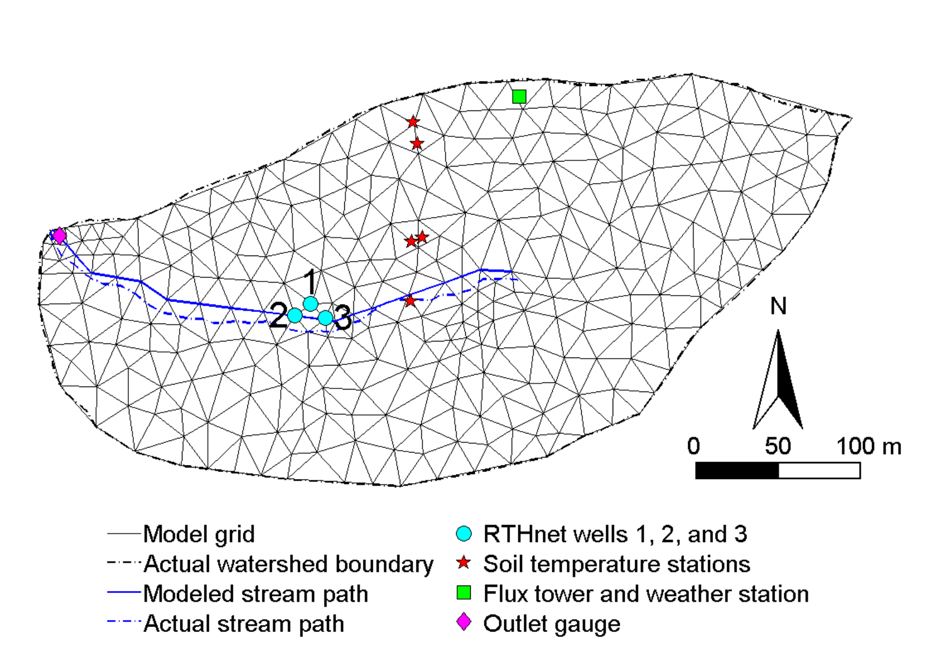
Parameter Estimation of Physically-Based Distributed Land Surface Hydrologic Model Using Ensemble Kalman Filter



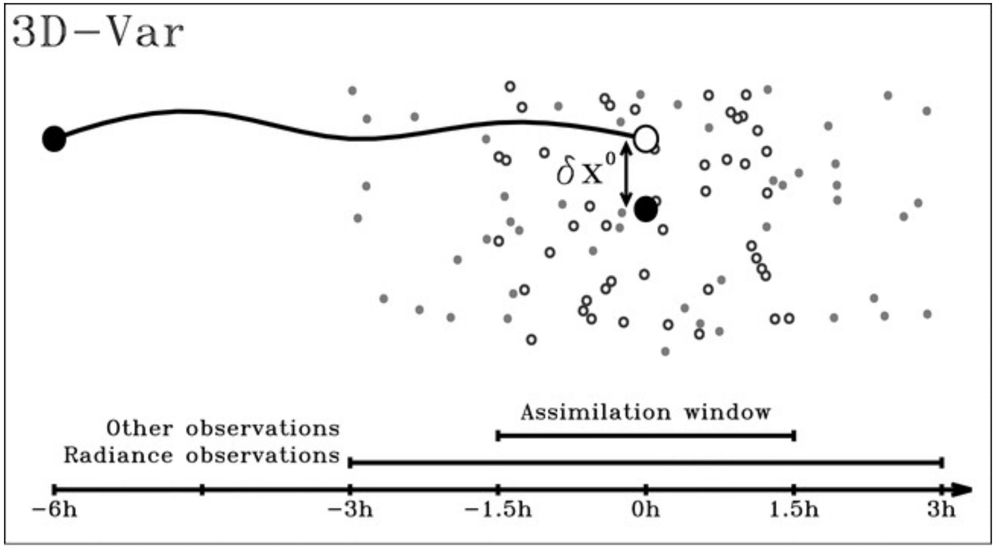
Flux-PIHM

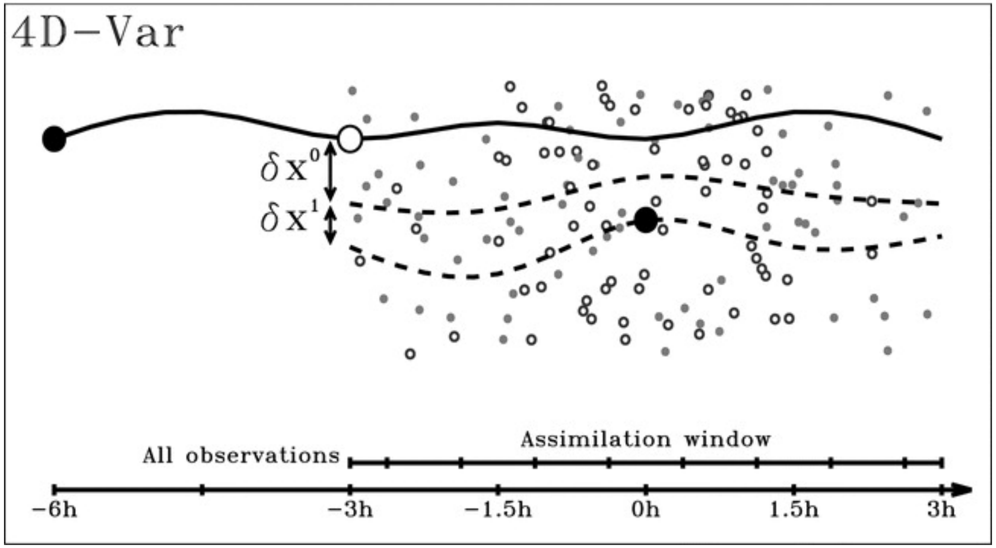
- *Fully-coupled land-surface hydrologic model*
- *Land-surface scheme is mainly adapted from Noah LSM*
- *Fully coupled surface water, soil water, groundwater, and land surface components*

- Observations
  - Discharge
  - Groundwater level and soil moisture
  - Land surface temperature
  - Latent heat flux



Advantages of 4DVar over 3DVar





i) Dealing with asynchronous observations

ii) Obtaining implicit flow-dependent background error covariance through linear models

iii) Using a forecast model as a dynamic constraint, enhancing the dynamic balance of the analysis

Data Assimilation: Terminology

The process to obtain a posterior estimate or analysis at a given time is often called data assimilation in atmospheric sciences.

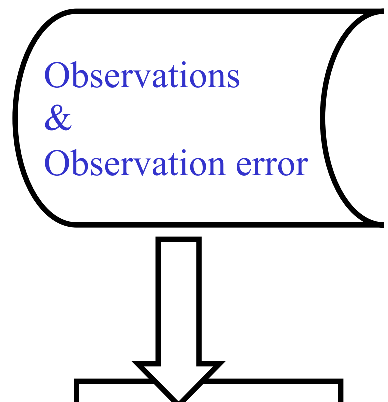
It usually combined observations with short-term forecast

x: State, dimension ~10^6-10^8

y: Observation, dimension ~10^3-10^6

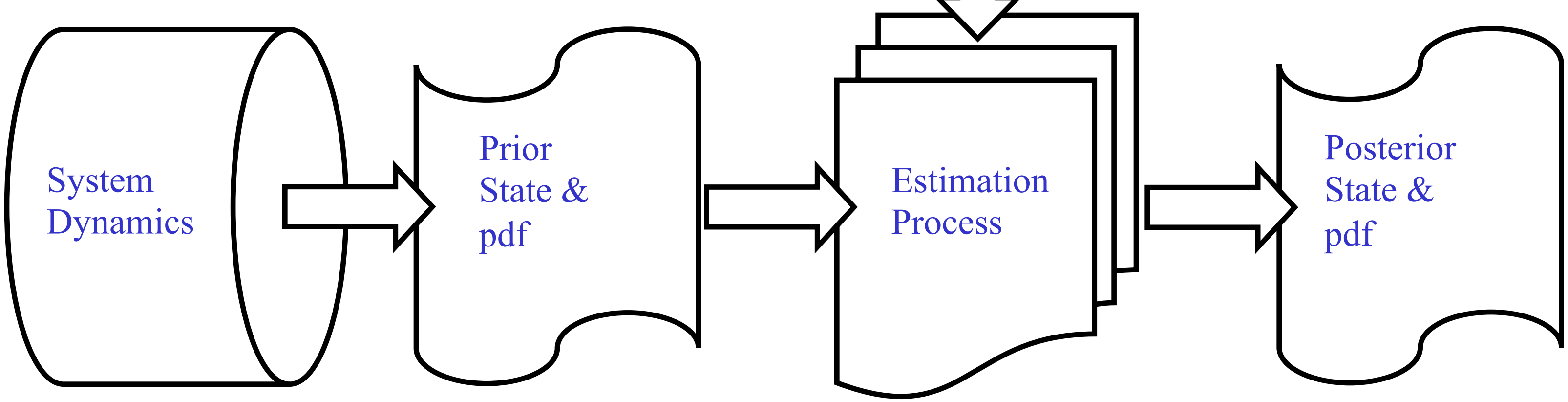
H: Observation operator

M: Dynamical model operator



$$\mathbf{y}_k^0 = \mathbf{H}\mathbf{x}_k^t + \varepsilon_k$$

$$(\mathbf{y}_k^0, \mathbf{R})$$



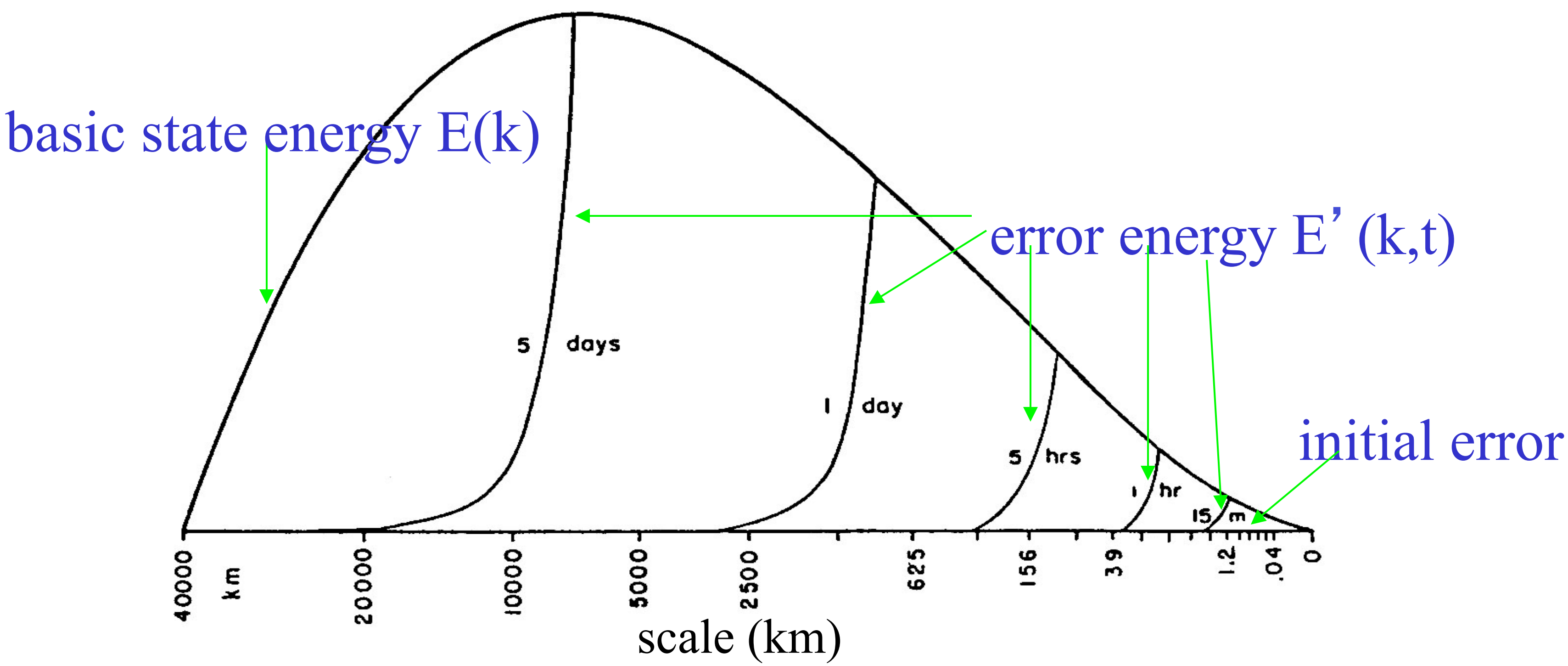
$$\mathbf{x}_k^t = \mathbf{M}\mathbf{x}_{k-1}^t + \eta_{k-1}$$

$$\mathbf{x}_k^f = \mathbf{M}\mathbf{x}_{k-1}^a$$

$$(\mathbf{x}_k^f, \mathbf{P}_k^f)$$

$$(\mathbf{x}_k^a, \mathbf{P}_k^a)$$

Multi-Scale Predictability Foreseen by Lorenz (1969)



"An error in observing a thunderstorm, after doubling perhaps every fifteen minutes until it becomes large, may subsequently lead to an error in a larger scale of motion, which may then proceed to double every five days. If this is the case, cutting the original error in half would increase the range of predictability of the larger scale not by five days but by only fifteen minutes."

Growing Trends and Challenges

- Coupling: air-ocean-land-chemistry-ecosytem-human behavior
- High resolution: global nonhydrostatic cloud-resolving model (1km)
- Limit of predictability: search for flow and scale dependent predictability given IC/model error and nonlinear/chaotic nature
- Use of ensembles: for uncertainty and probabilistic estimates using multi-model, stochastic physics, with IC uncertainty
- Observational network design and observation targeting: how to best observe, how to quality control and bias correction, …
- Advanced data assimilation: from 3DVAR to 4DVAR or EnKF, to hybrid and coupling of EnKF, 3DVAR and 4DVAR
- Simultaneous state and parameter estimation for better understand the physics and treatment of model error
- Interdisciplinary: stats, applied math, computer science, social science, economics, decision making, data mining, cyberscience, …
- Tech Enabling: academics and educator for quasi-operational NWP

Concluding Remarks on Future Atmospheric Prediction

- Better understanding of under-resolved, under-observed, less-predictable processes and their impacts on predictability
- Improve models through better physics and finer resolution
- Design and improve observing network, such as hurricane inner-core observations from airborne Doppler radars
- Implement and develop advanced data assimilation techniques to better utilize existing and future observations
- Going probabilistic: ensemble-based initialization and forecasting
- Design computationally efficient numerics for model integration and data assimilation
- Combine imperfect model/data for better understanding physics (PE)
- Better data sharing, communication, visualization
- Giant computers like the TACC Ranger cluster