Some of our Recent Research on Ensemble-based Data Assimilation

Fuqing Zhang

Penn State University

Sponsored by NASA, ONR, NSF and NOAA

REVIEW

Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation

P. L. HOUTEKAMER

Meteorology Research Division, Environment and Climate Change Canada, Dorval, Québec, Canada

FUQING ZHANG

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 17 December 2015, in final form 6 June 2016)

ABSTRACT

This paper reviews the development of the ensemble Kalman filter (EnKF) for atmospheric data assimilation. Particular attention is devoted to recent advances and current challenges. The distinguishing properties of three well-established variations of the EnKF algorithm are first discussed. Given the limited size of the ensemble and the unavoidable existence of errors whose origin is unknown (i.e., system error), various approaches to localizing the impact of observations and to accounting for these errors have been proposed. However, challenges remain; for example, with regard to localization of multiscale phenomena (both in time and space). For the EnKF in general, but higher-resolution applications in particular, it is desirable to use a short assimilation window. This motivates a focus on approaches for maintaining balance during the EnKF update. Also discussed are limited-area EnKF systems, in particular with regard to the assimilation of radar data and applications to tracking severe storms and tropical cyclones. It seems that relatively less attention has been paid to optimizing EnKF assimilation of satellite radiance observations, the growing volume of which has been instrumental in improving global weather predictions. There is also a tendency at various centers to investigate and implement hybrid systems that take advantage of both the ensemble and the variational data assimilation approaches; this poses additional challenges and it is not clear how it will evolve. It is concluded that, despite more than 10 years of operational experience, there are still many unresolved issues that could benefit from further research.

CONTENTS

1. Introduction 4490
2. Popular flavors of the EnKF algorithm 4491
a. General description 4491
b. Stochastic and deterministic filters 4492
1) The stochastic filter 4492
2) The deterministic filter 4492
c. Sequential or local filters 4493
1) Sequential ensemble Kalman filters 4493
2) The local ensemble transform Kalman filter 4494
d. Extended state vector 4494
e. Issues for the development of algorithms 4495
3. Use of small ensembles 4495
a. Monte Carlo methods 4495
b. Validation of reliability 4497
c. Use of group filters with no inbreeding 4498
d. Sampling error due to limited ensemble size: The rank problem 4498
e. Covariance localization 4499
1) Localization in the sequential filter 4499
2) Localization in the LETKF 4499
3) Issues with localization 4500
f. Summary 4501
4. Methods to increase ensemble spread 4501
a. Covariance inflation 4501
1) Additive inflation 4501
2) Multiplicative inflation 4502
3) Relaxation to prior ensemble information 4502
4) Issues with inflation 4503
b. Diffusion and truncation 4503
c. Error in physical parameterizations 4504
1) Physical tendency perturbations 4504
2) Multimodel, multiphysics, and multiparameter approaches 4505
3) Future directions 4505
d. Realism of error sources 4506

Houtekamer & Zhang (December 2016, MWR) #2 MWR most-read (1787 downloads)

PSU WRF-based multi-functional regional-scale ensemble and hybrid data assimilation system

DA methods included:

PSU **WRF-EnKF** (Zhang et al. 2009a; Weng & Zhang 2012): publically released

NCAR WRFDA-**3DVar** (Huang et al. 2009): publically released

NCAR WRFDA-**4DVar** (X Zhang et al. 2014): publically released

E3DVar/3DenVar (hybrid/coupling of EnKF & 3DVar) (Zhang et al. 2013)

E4DVar (coupling of EnKF & 4DVar) (Zhang et al. 2009b; Zhang & Zhang 2012)

4DEnVar (ensemble-based 4D hybrid) (Liu et al. 2008; Poterjoy & Zhang 2016)

Current DA plans at the leading NWP centers:

ECWMF: adjoint-based as an ensemble of 4DVar but with hybrid covariance

UK-Met: adjoint-based E4DVar in operation, better than ensemble-based 4DEnVar

NCEP: ensemble-based 4DEnVar

CMC: 4DEnVar for deterministic forecasts, EnKF for ensemble prediction

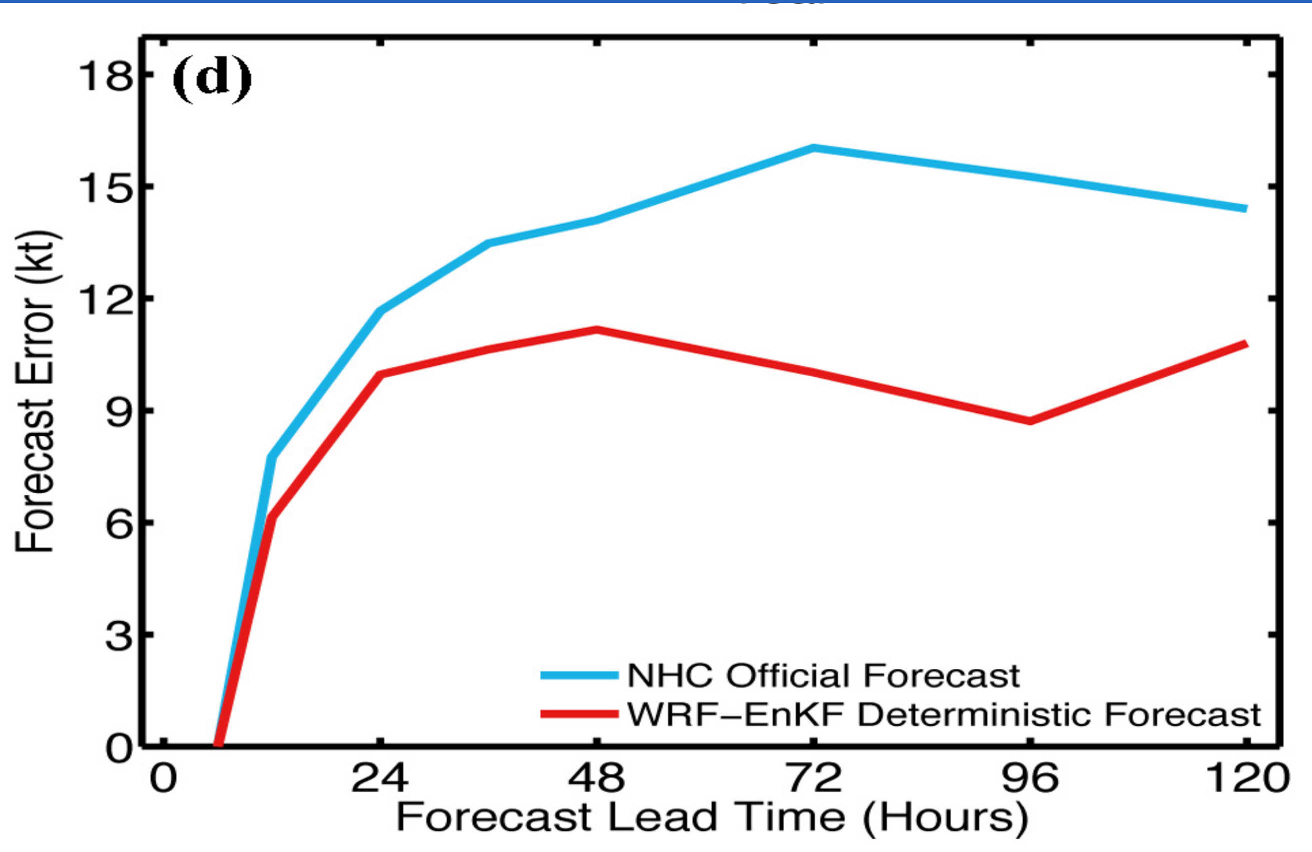
WRF-EnKF Performance Assimilating Airborne Vr

all 100+ P3 TDR missions during 2008-2012

Quasi-operational evaluation by NOAA/NHC since 2011 as stream 1.5 run

WRF-EnKF: 3 domains (27, 9 , 3km), 60-member ensemble, PSU TC flux scheme

Intensity error (knots)



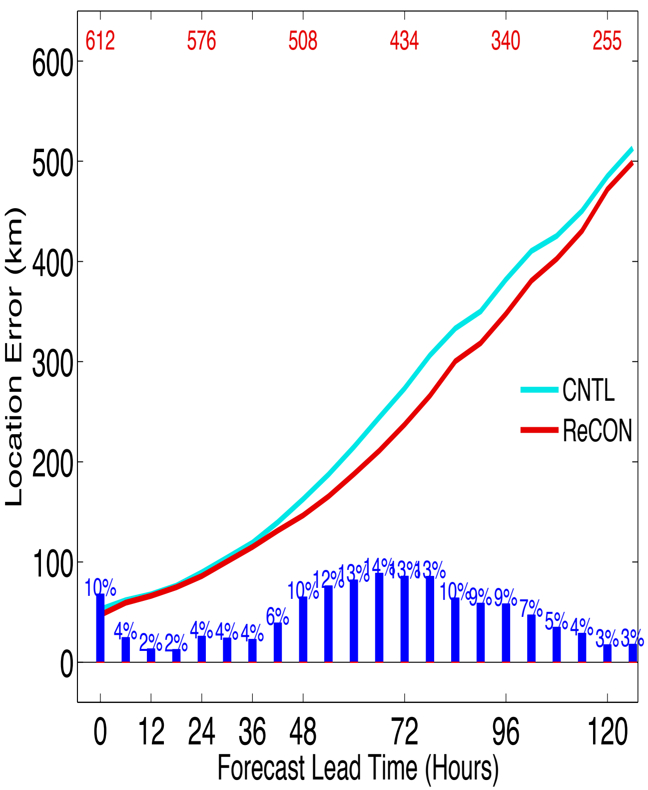
(Zhang et al. 2011 GRL; Zhang and Weng, 2015 BAMS)

WRF-EnKF Performance w/ versus w/o Aircraft OBS

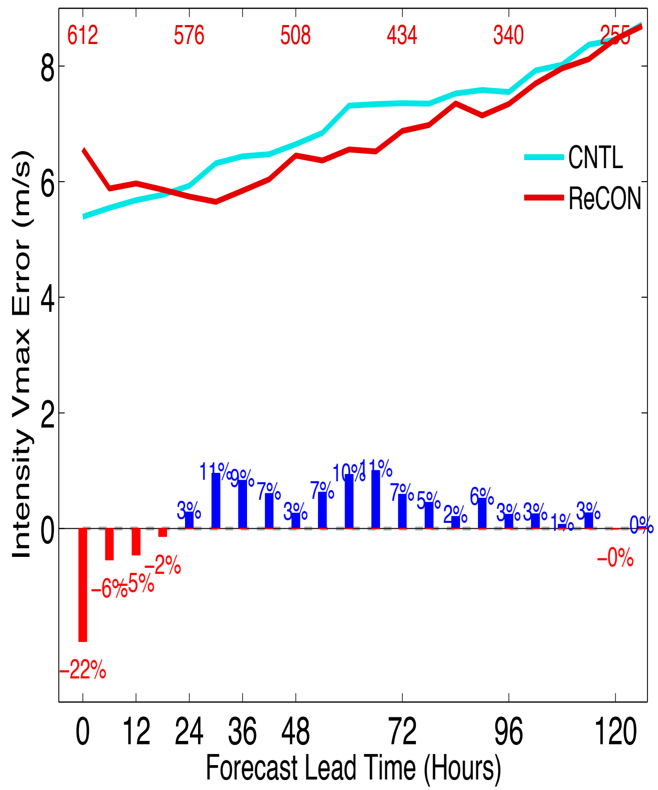
for HFIP/NHC selected RDITT cases w/o TDR during 2008-2012

WRF-EnKF: 3 domains (27, 9 , 3km), 60-member ensemble, PSU TC flux scheme

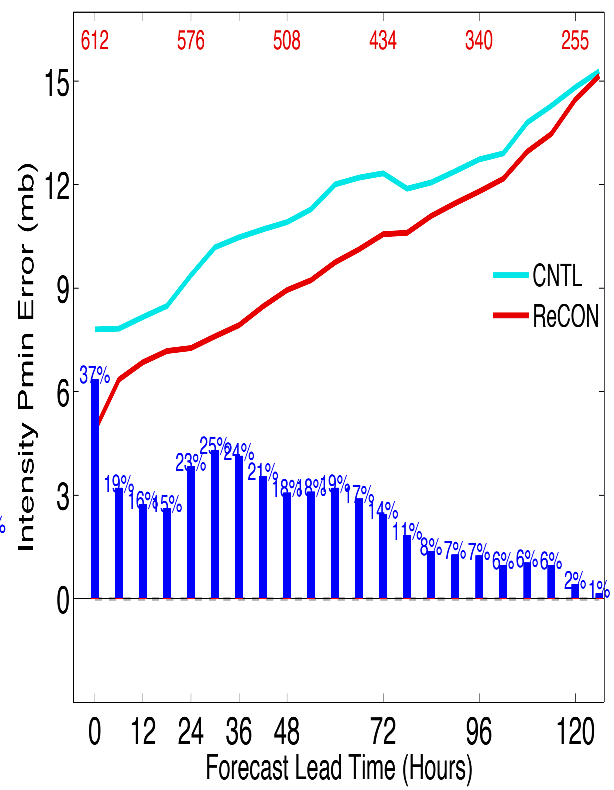
Position error (km)



Vmax error (knots)



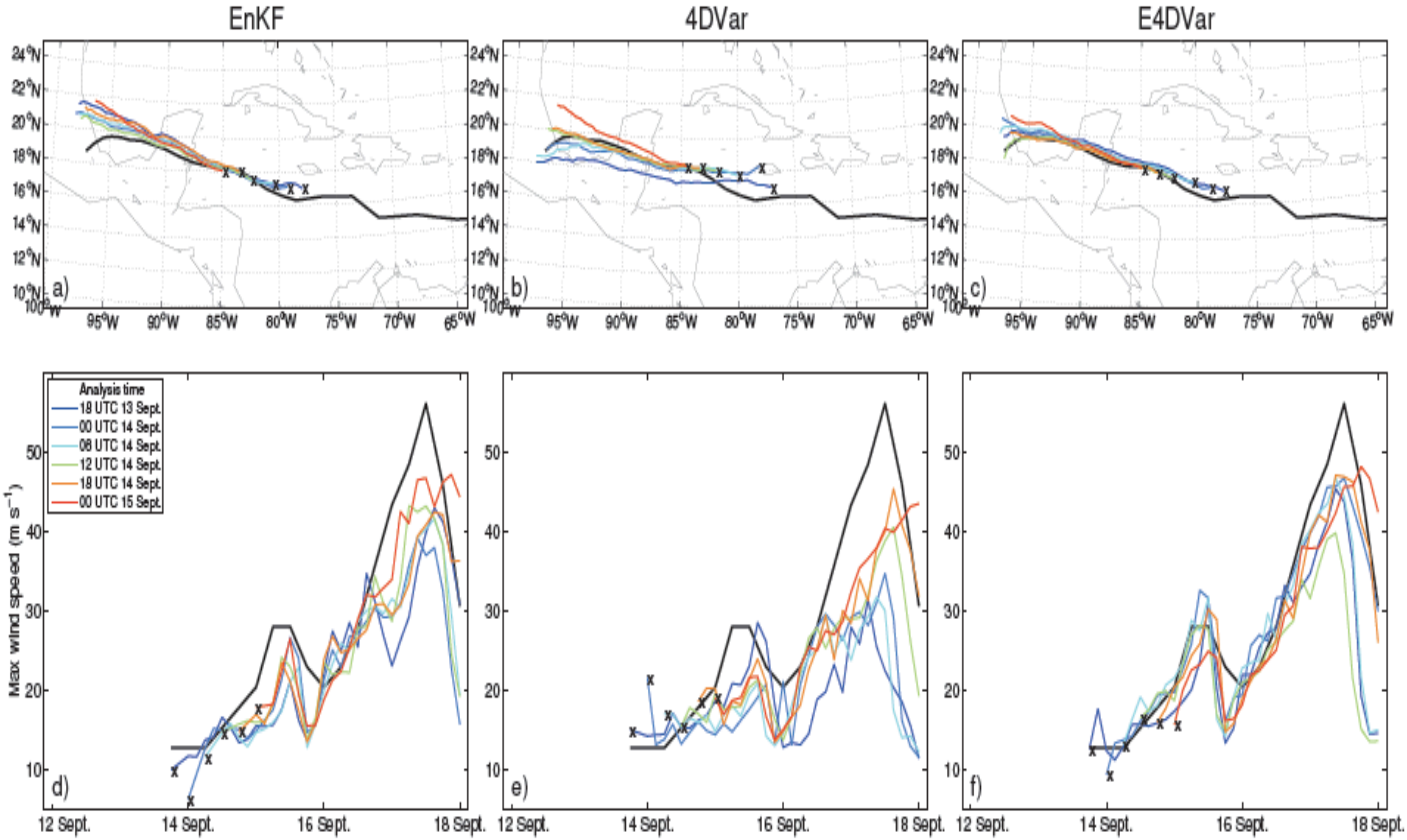
Pmin error (mb)



(Weng and Zhang, 2016 JMSJ)

Inter-comparison of E4DVar vs. EnKF & 4DVar for TCs

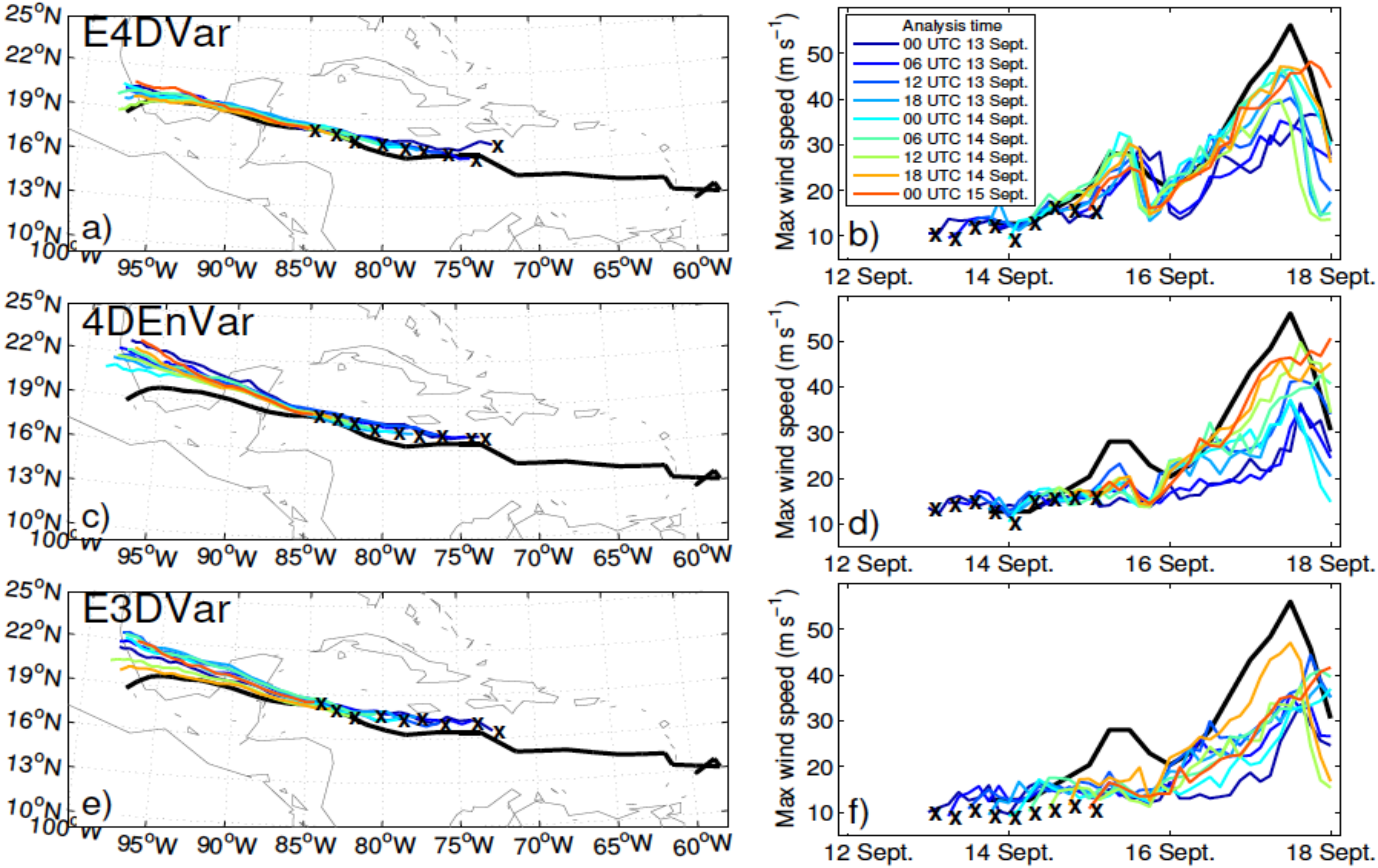
Deterministic forecast for Track & Intensity: w/ field sondes



(Poterjoy and Zhang, 2014 MWR)

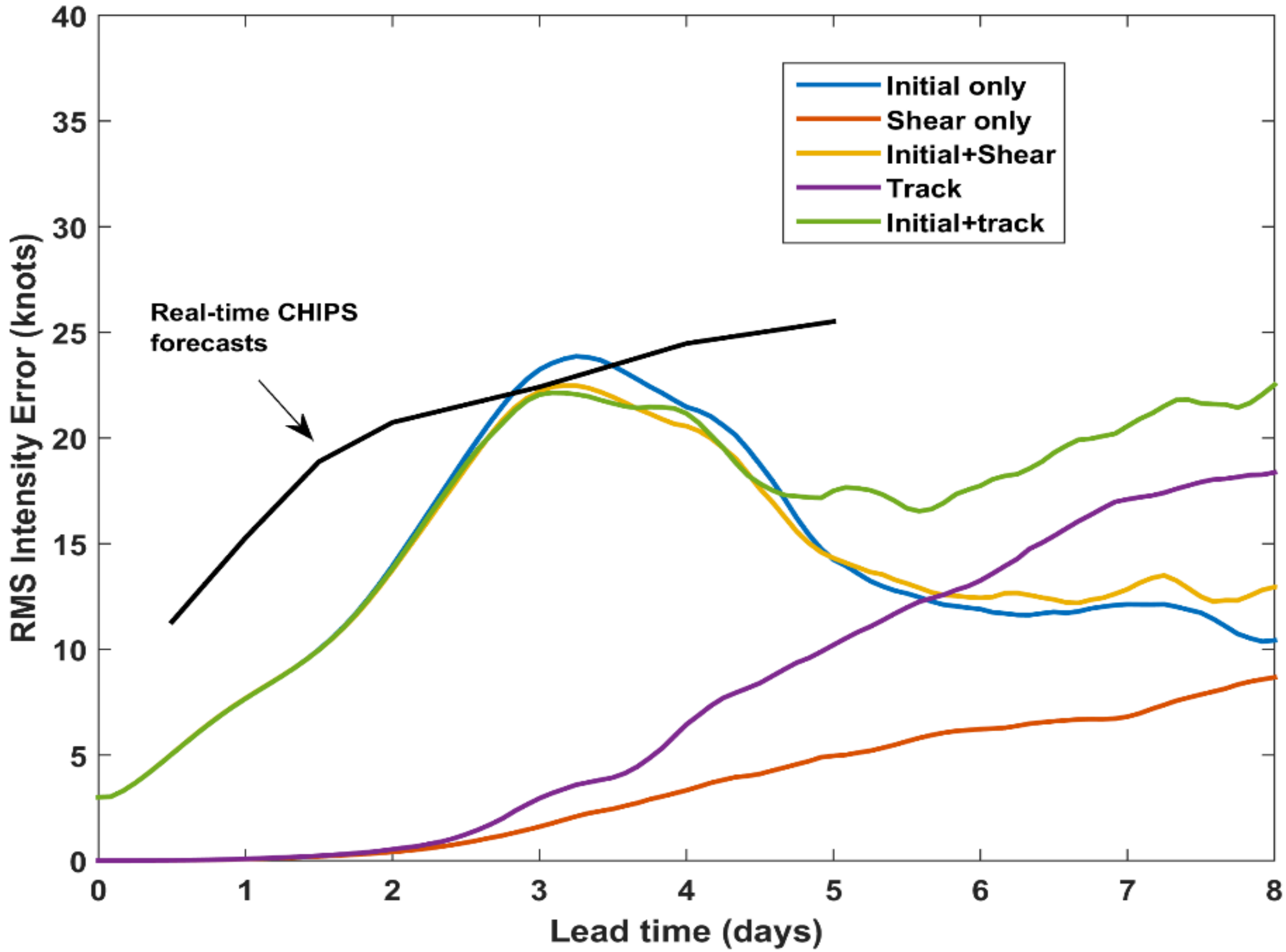
Inter-comparison of E4DVar vs. 4DEnVar and E3DVar

Deterministic forecast for Track & Intensity: w/ field sondes



(Poterjoy and Zhang, 2016 MWR)

Predictability and Error Sources of TC Intensity Forecasts: Lessons Learned from CHIPS 2009-2015



(Emanuel and Zhang 2016 JAS)

#2 JAS most-read (1125 downloads)

Hurricane Edouard (2014) Ensemble Track & Intensity

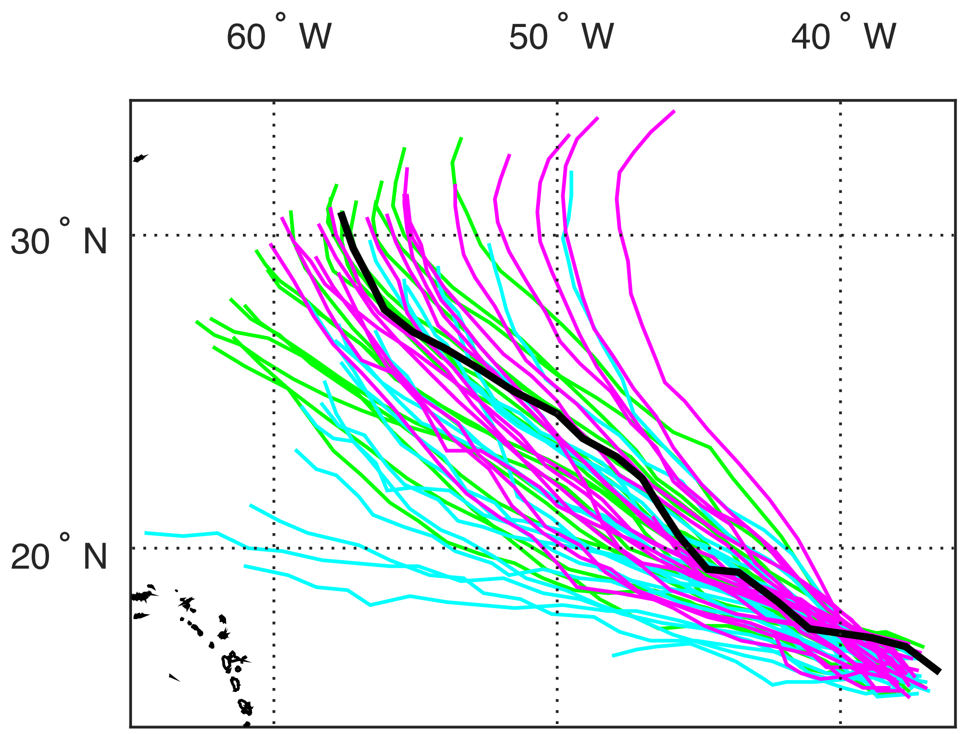
Multi-core ensemble spread larger than any individual single-core

5 day forecast initialized
2014-09-11 12 UTC

Spread:

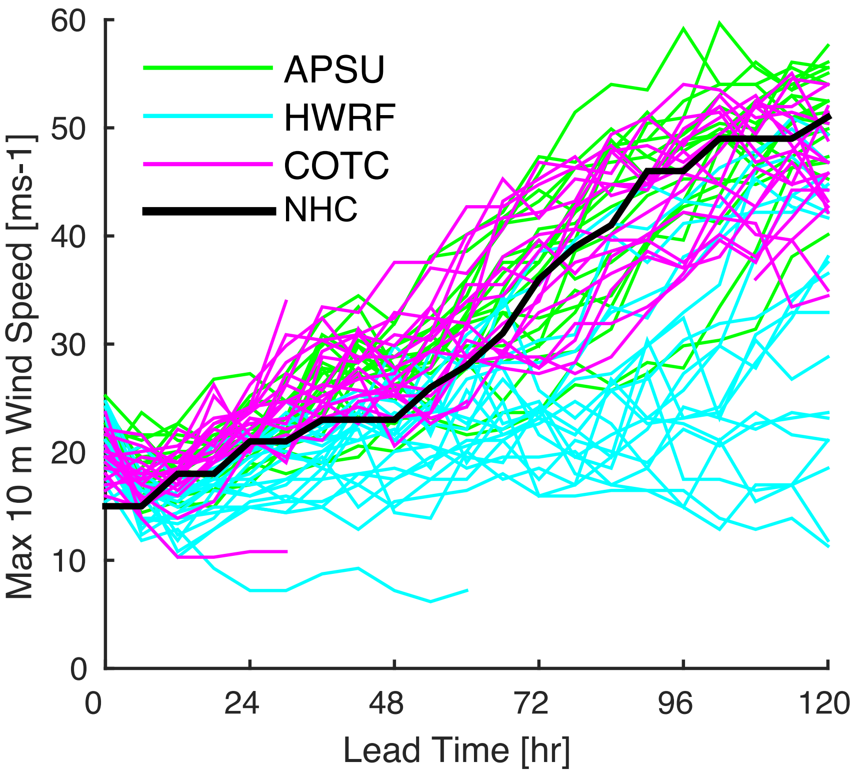
APSU ~ COTC !~ HWRF

Ensemble Track*



**20 of 60 members shown*

Ensemble Intensity*



(Melhauser, Zhang et al. 2016 WAF)

Hurricane Edouard (2014) Ensemble Mean - Physics

"APSU-Like" Physics

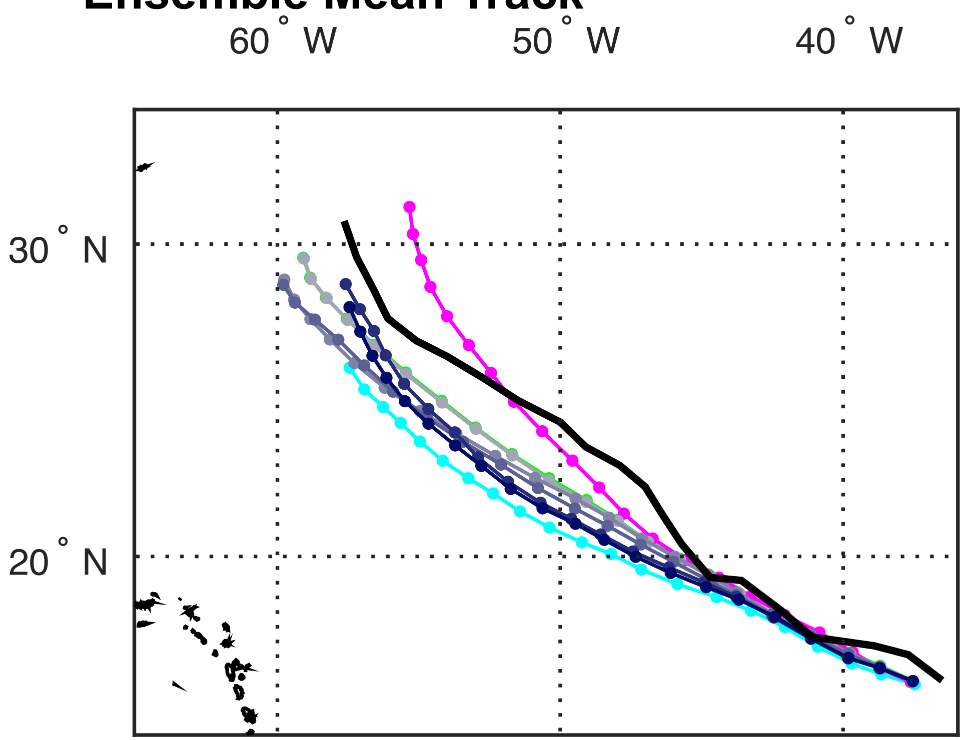
Modify microphysics, radiation, PBL, surface drag, cumulus

"HWRF-Like" Physics

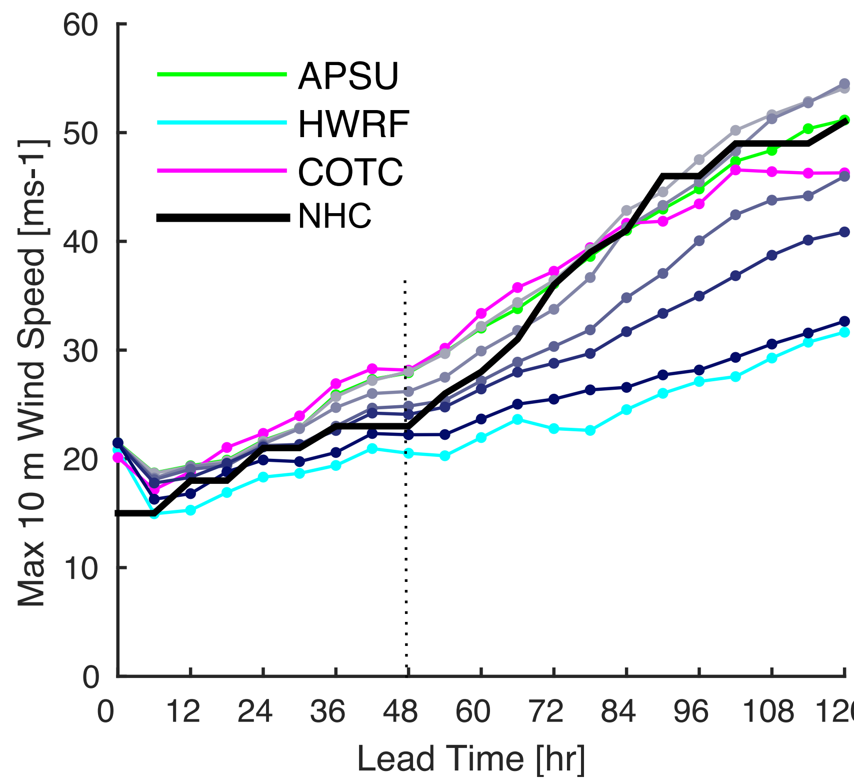
Shift single-core mean to behave similarly to a different model-core

- Physics configuration has a leading influence
- More evident in intensity

Ensemble Mean Track



Ensemble Mean Intensity



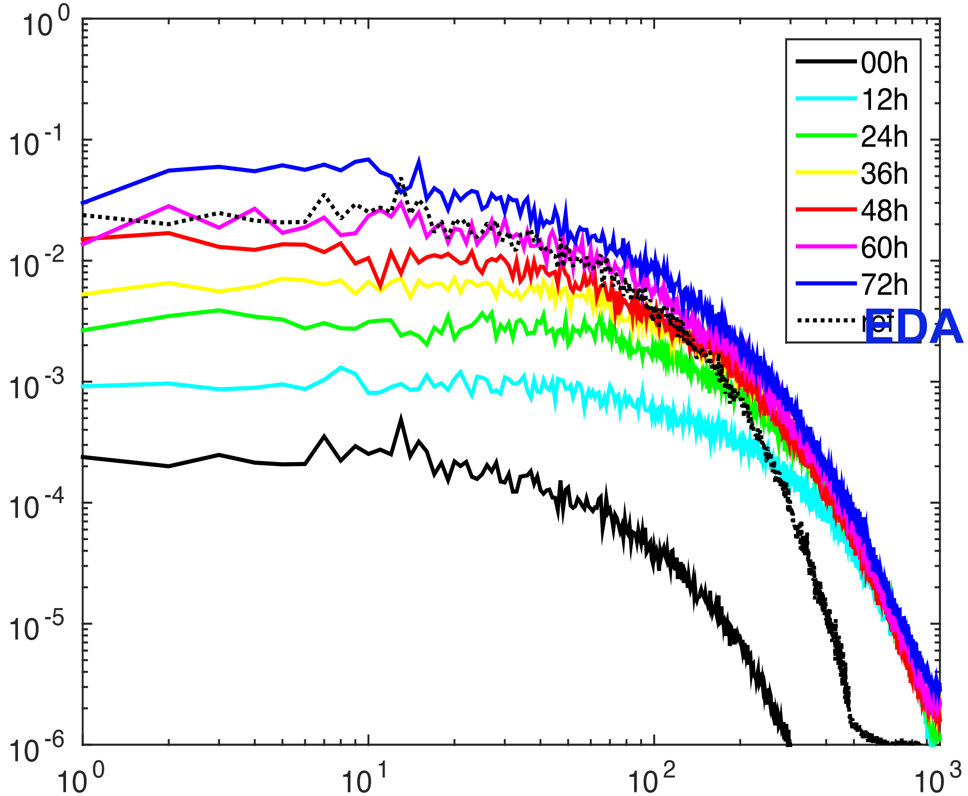
(Melhauser, Zhang et al. 2016 WAF)

What is the Ultimate Limit of Midlatitude Weather Predictability?

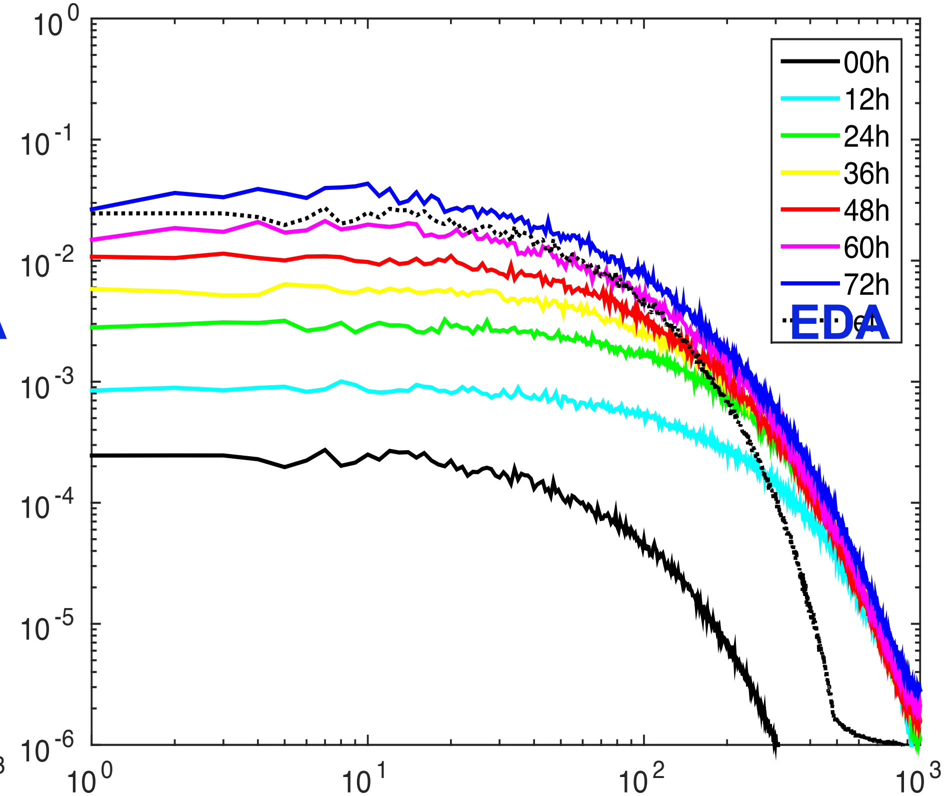
It takes about 3 days for 10%EDA IC error run to grow to 100%EDA!

It takes about 1 days for 70%EDA IC error run to grow to 100%EDA

Error spectra of ECMWF IFS member 1



Error spectra of 10-member mean

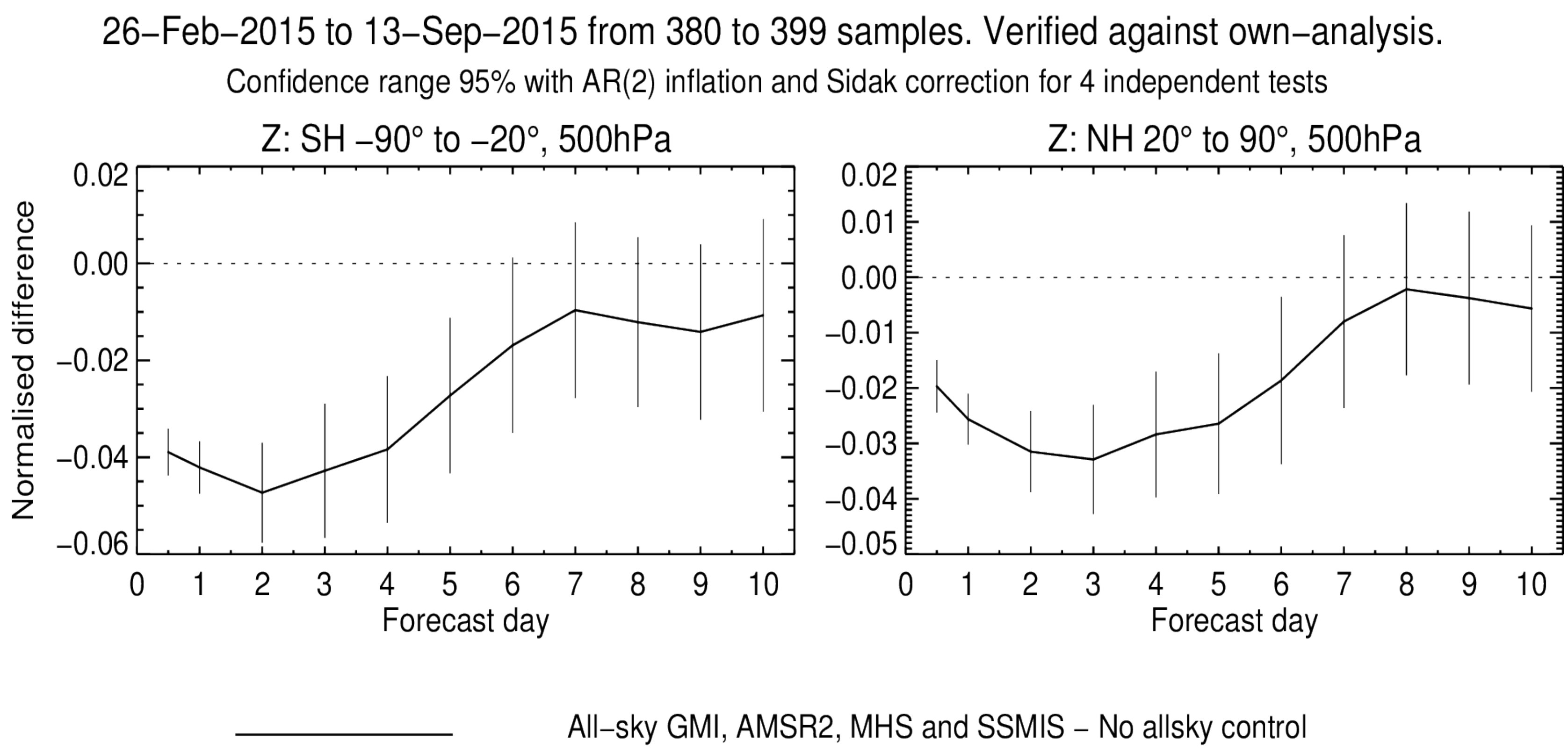


Ongoing collaboration with Linus Magnusson and Roberto Buizza at ECMWF, Y.Q. Sun at PSU

State-of-the-Science: Importance of Cloudy and Precipitating Scenes

High FSO => real improvements in medium-range synoptic forecasts

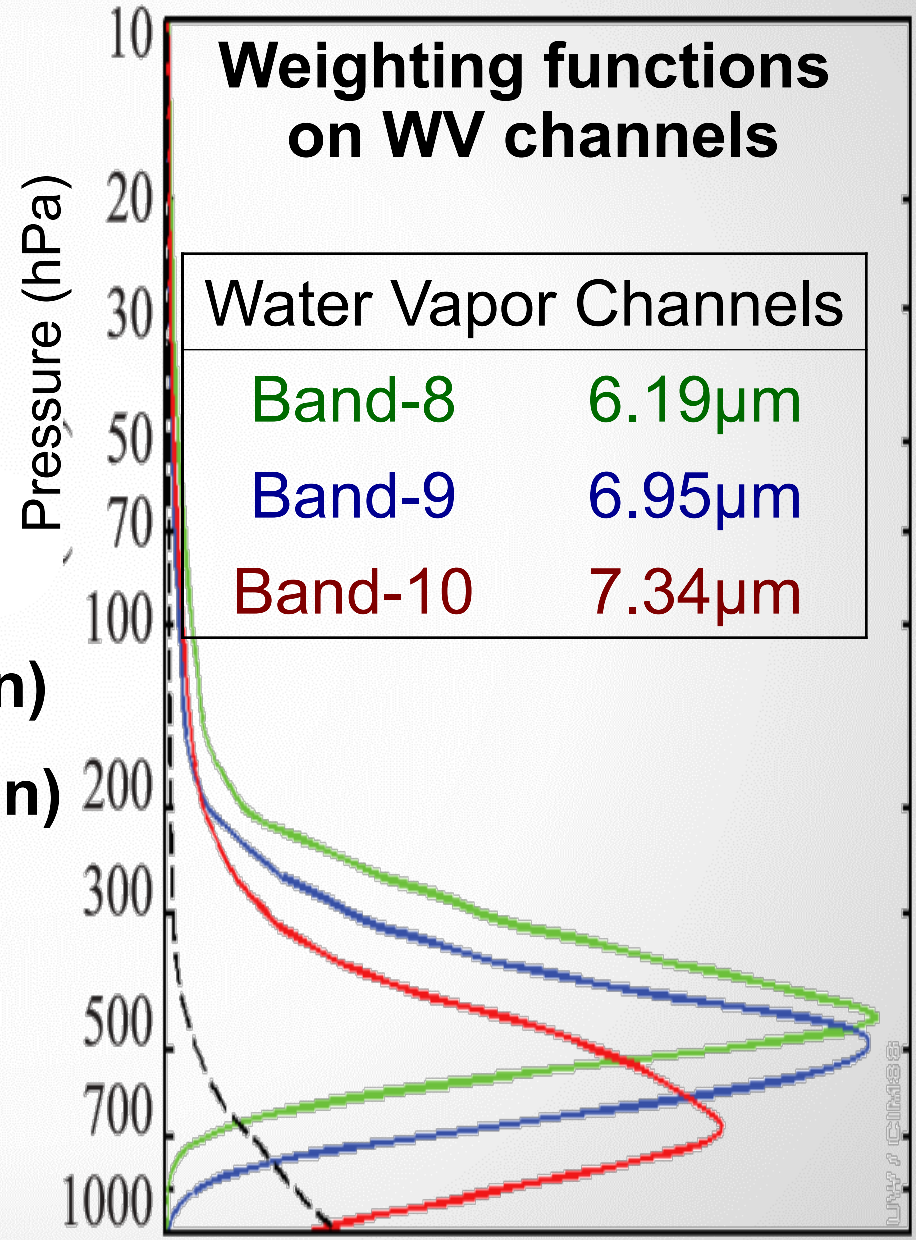
Mechanism: *4D-Var can infer dynamical initial conditions from observed WV, cloud and precipitation*



ECMWF

Courtesy of Alan Geer at ECMWF

New Generation of Geostationary IR Satellites



Launch Dates:

- **Oct 2014 (Himawari-8, Japan)**
- **Nov 2016 (Himawari-8, Japan)**
- **Nov 2016 (GOES-R, USA)**
- **Dec 2016 (FY-4, China)**

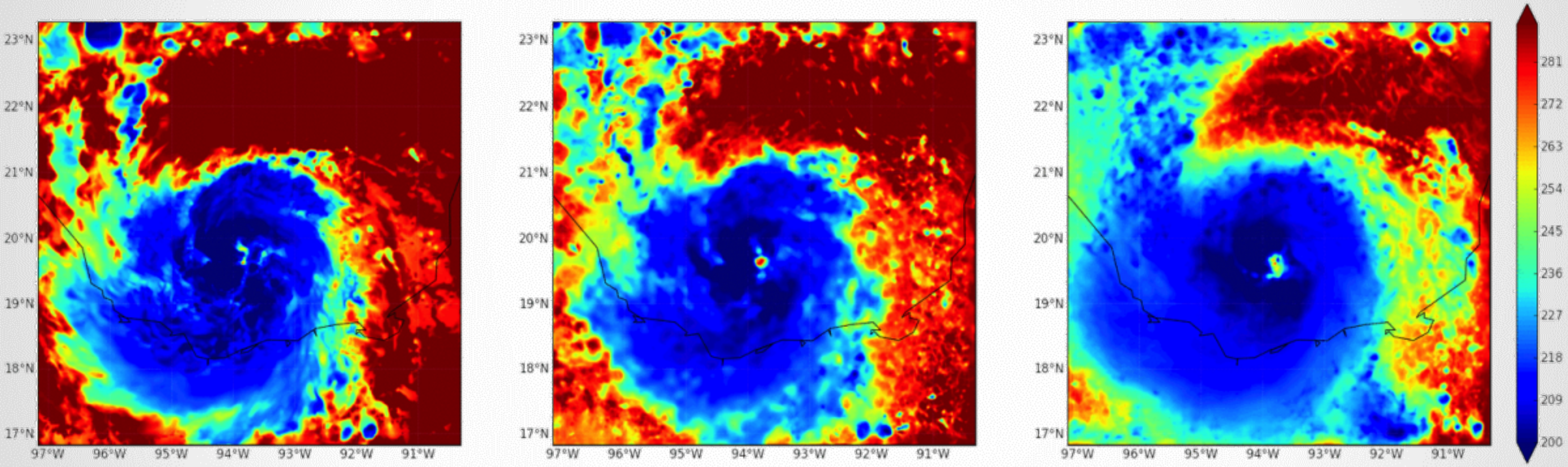
High spatiotemporal resolutions:

➢ **10-15 minutes; 2-4 km**

EnKF Performance assimilating simulated radiance

Truth versus EnKF-analyzed Infrared Radiance
of GOES-R ABI ch14 (11.2 µm)

[2010-09-16_22:00]



Verifying truth

EnKF analysis
with radiance &
minimum SLP

EnKF analysis
with minimum
SLP only

(Zhang, Minamide & Clothiaux, 2016 GRL)

Adaptive Observation Error Inflation (AOEI)

Problem: **erroneous analysis increments**

If **Model (clear / cloudy) ≠ Observation (cloudy / clear)**

In updating SLP, $\frac{12.5\,[hPa \times K]}{3^2 + 5^2 [K^2]} \times 40[K] \sim \mathbf{15[hPa]}$

AOEI: inflating observation error variance

$$\sigma^2_{o-AOEI} = max \left\{ \sigma^2_o, [y_o - h(x_b)]^2 - \sigma^2_{h(x_b)} \right\}$$

With AOEI, $\frac{12.5\,[hPa \times K]}{40^2 [K^2]} \times 40[K] \sim \mathbf{0.3[hPa]}$

AOEI

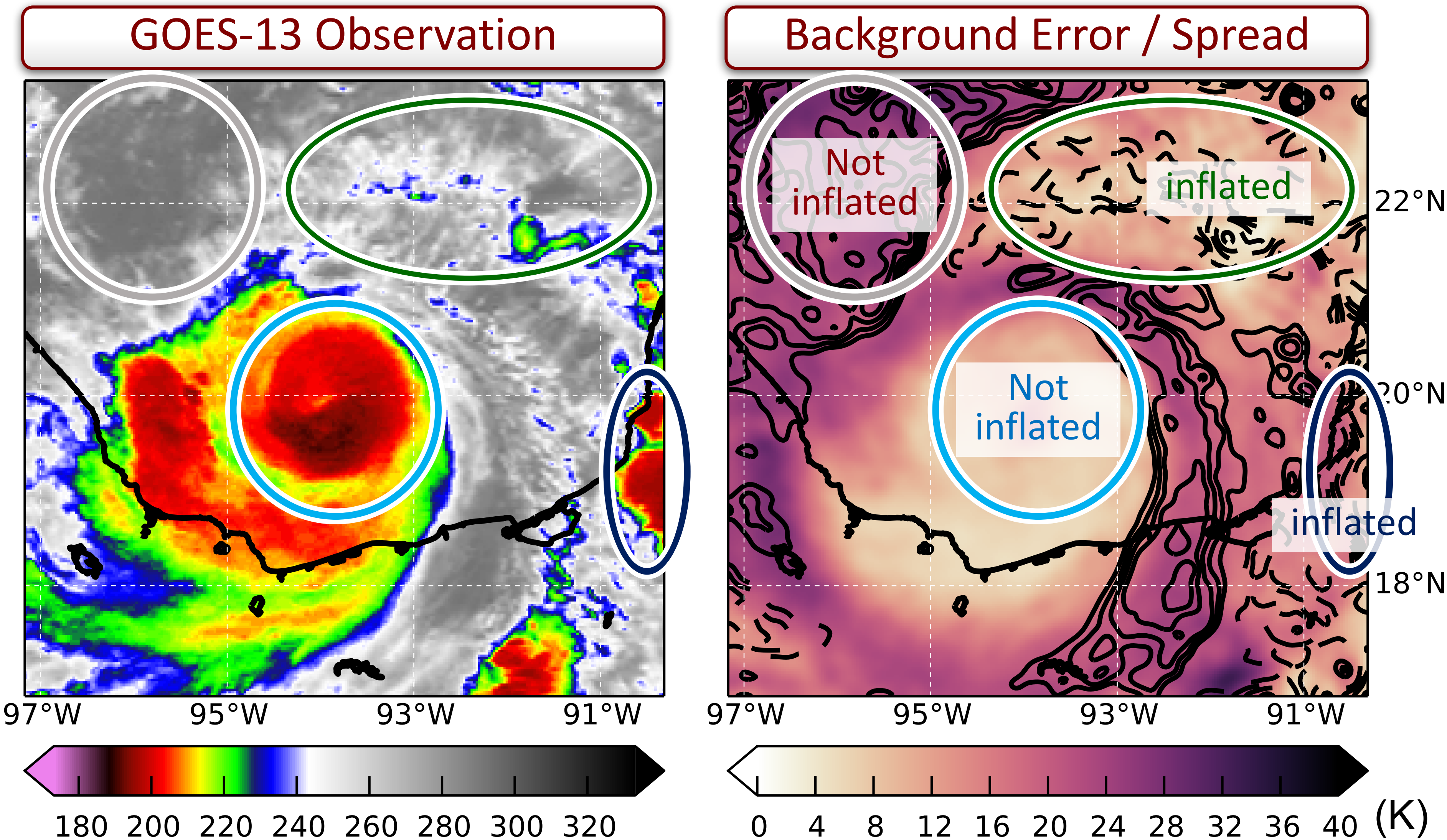
suppresses erroneous analysis increments,

relieves the issues of representativeness & sampling,

& contributes to maintaining balance.

(Minamide & Zhang, MWR, 2017)

Adaptive Observation Error Inflation (AOEI)

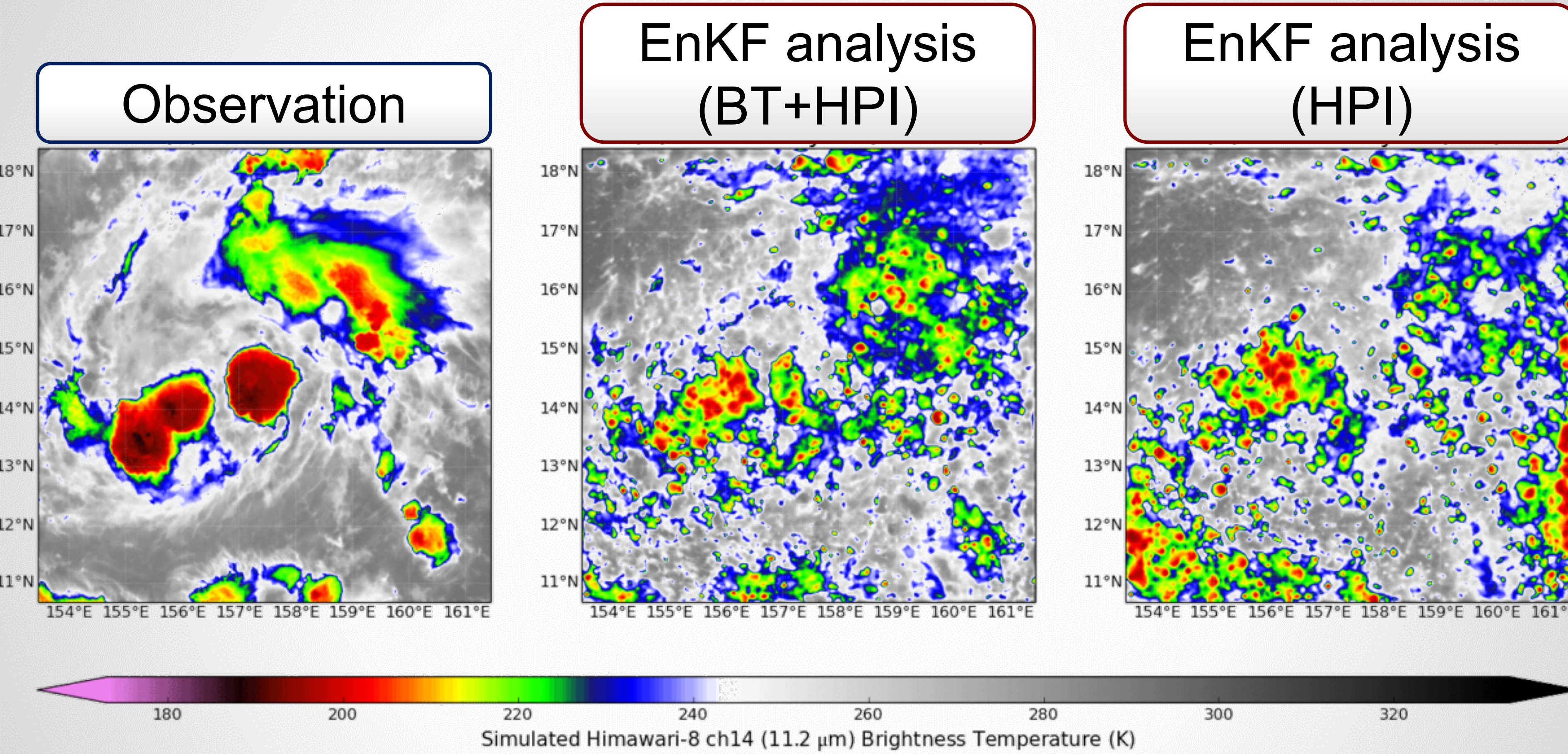


Color: observed brightness temperature

Contour: Background error
Color: ensemble spread

EnKF Performance W/ Assimilating Himawari-8 BT

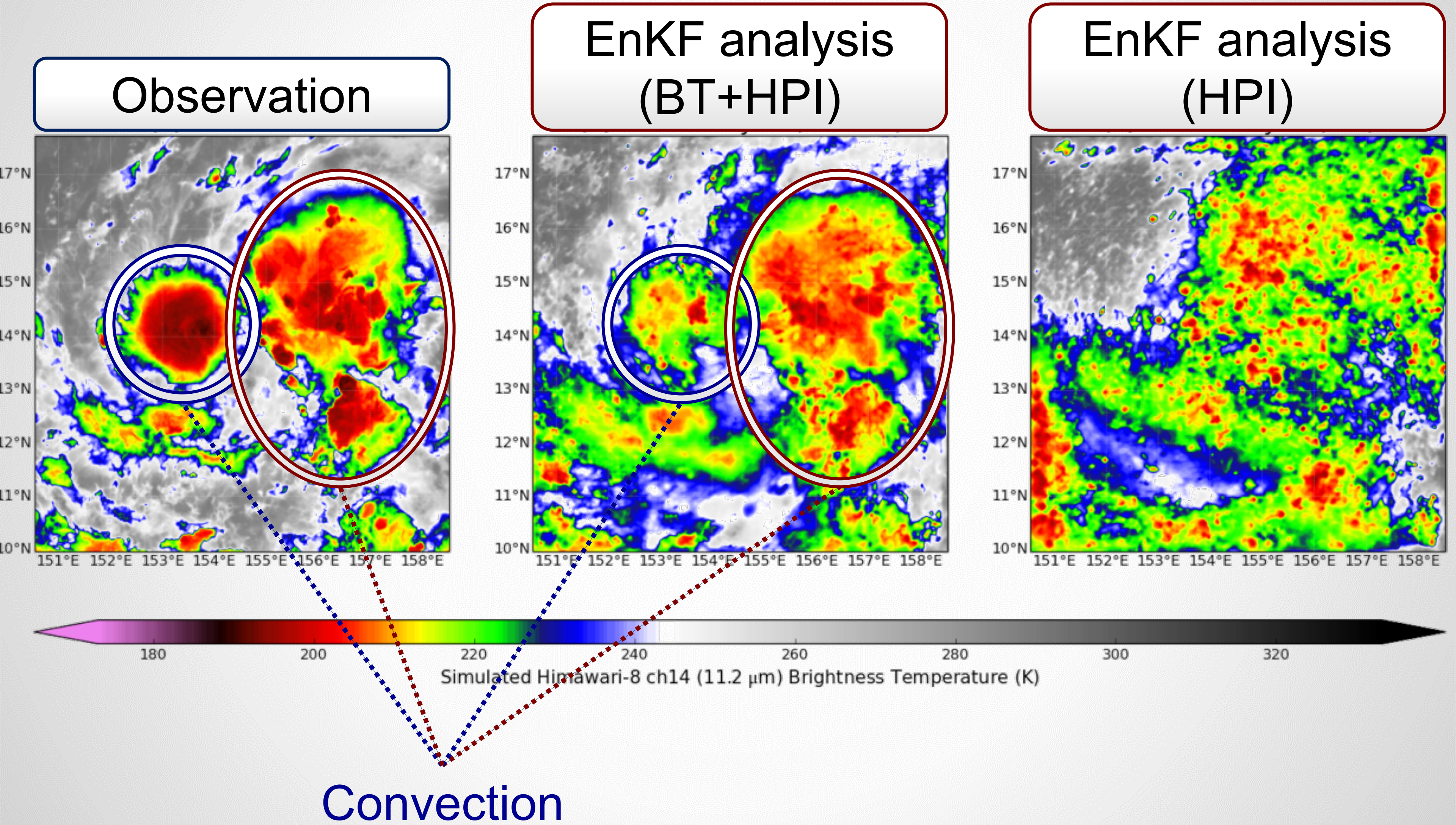
Himawari-8 **Infrared** Channel (ch14: 11.2 μm)



On going work with Masashi Minamide

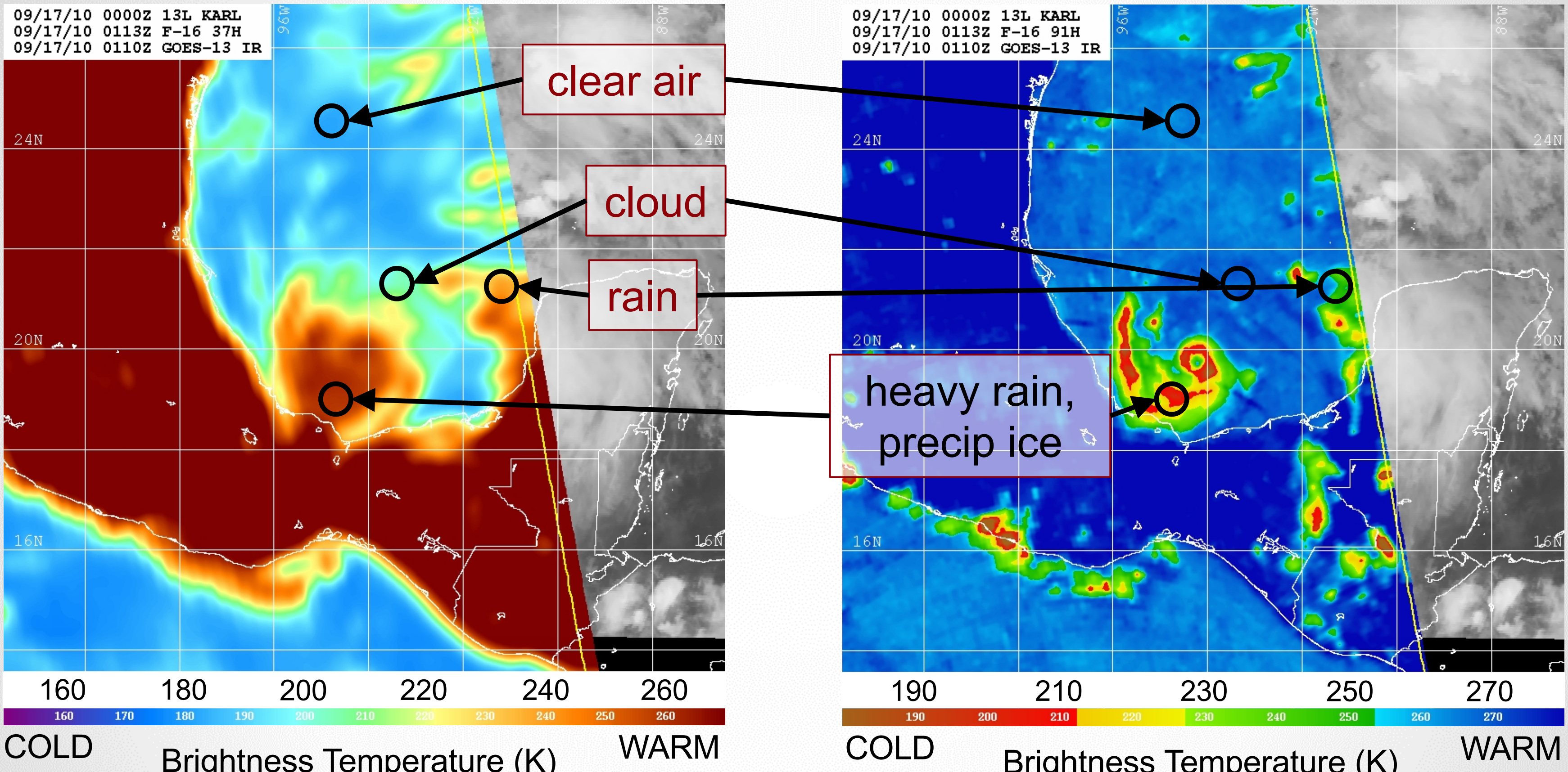
EnKF Performance W/ Assimilating Himawari-8 BT

Himawari-8 **Infrared** Channel (ch14: 11.2 μm)



Microwave Radiometers and Precipitation

Low-mid microwave freq. (37 GHz) High-mid microwave freq. (91.7 GHz)

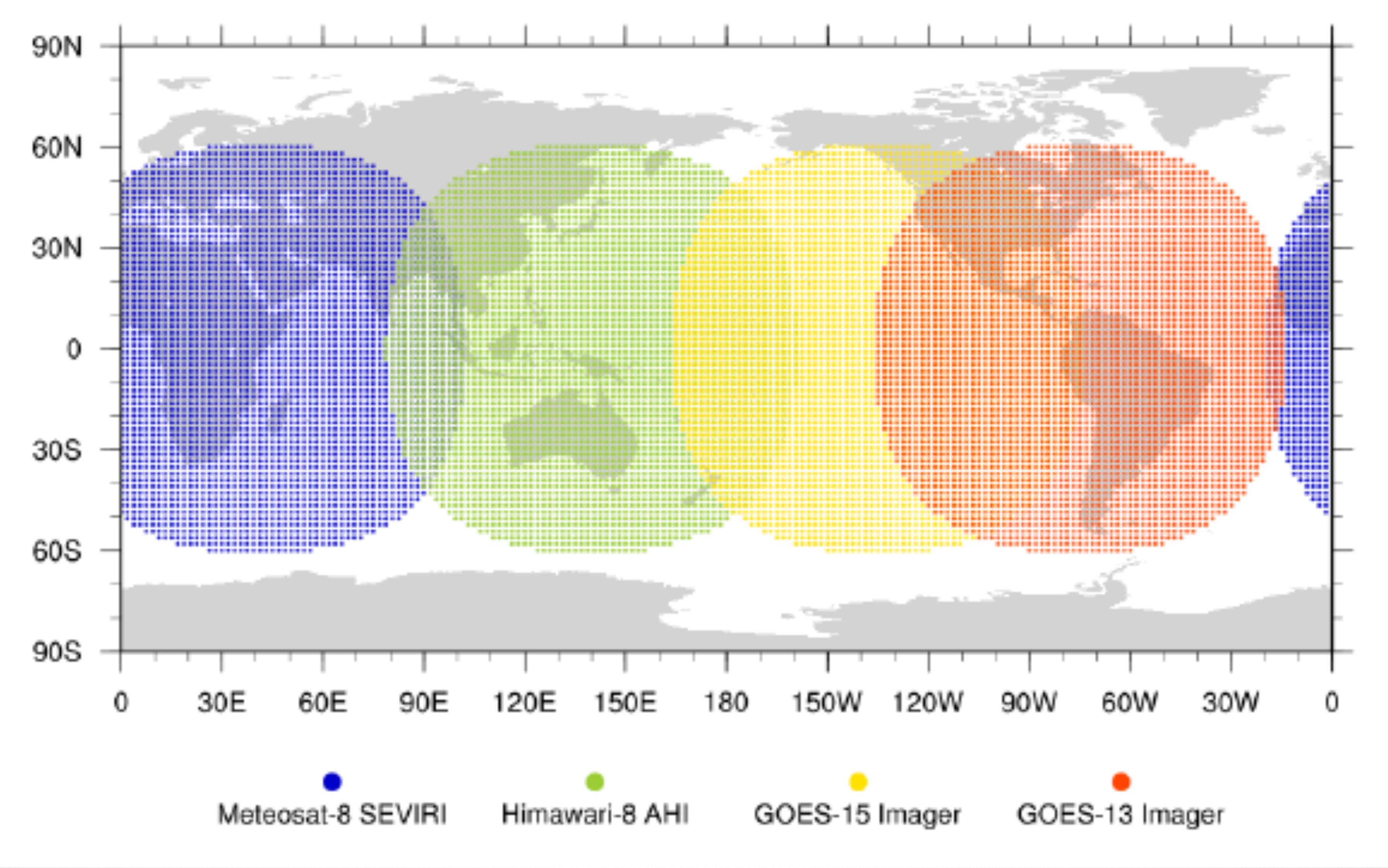


Hurricane Karl 09/17/10 0113Z
(SSMI/S image courtesy NRL)

- Rain and cloud liquid net add to low emission by water
- Some scattering by precipitation ice
- Scattering by precipitation ice dominates the signal

On going work with Scott Sieron, Eugene Clothiaux and Yinghui Lu

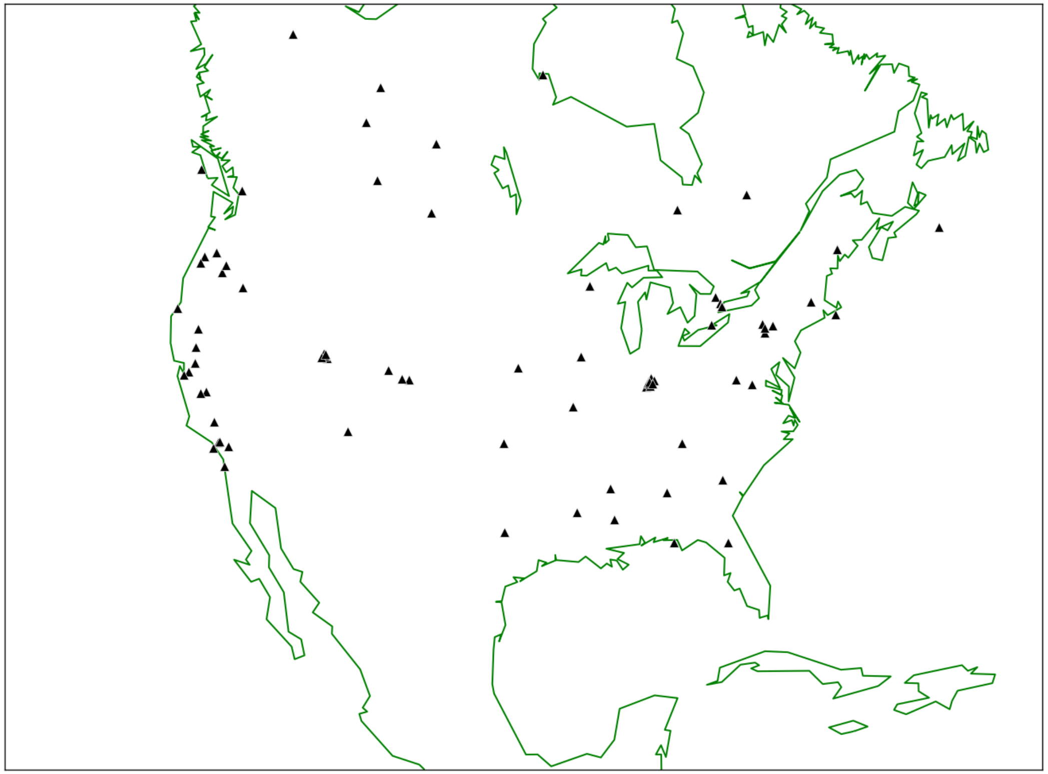
Global IR coverage & ongoing GFS/GSI-LETKF OSSE



Ongoing collaboration with Da Cheng and Eugenia Kalnay at UMD

PSU WRF-Chem-based EnKF carbon DA system

In the initial OSSE experiment, we assimilated the truth at the following tower locations every 1 hour

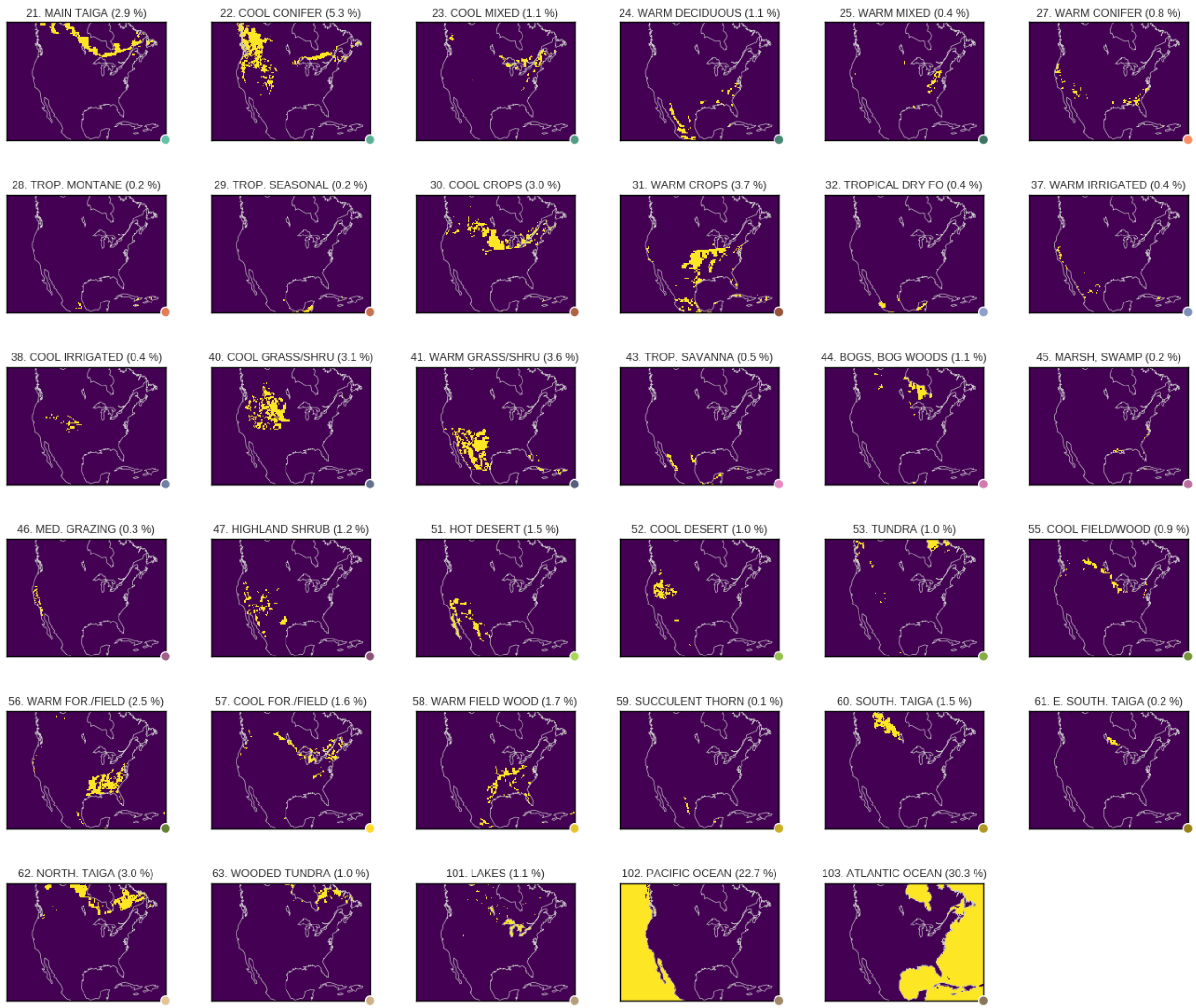


Ongoing work with graduate student Hans Chen

35 ecosystems based on the Olson (1992) classification

One α parameter is estimated for each ecosystem

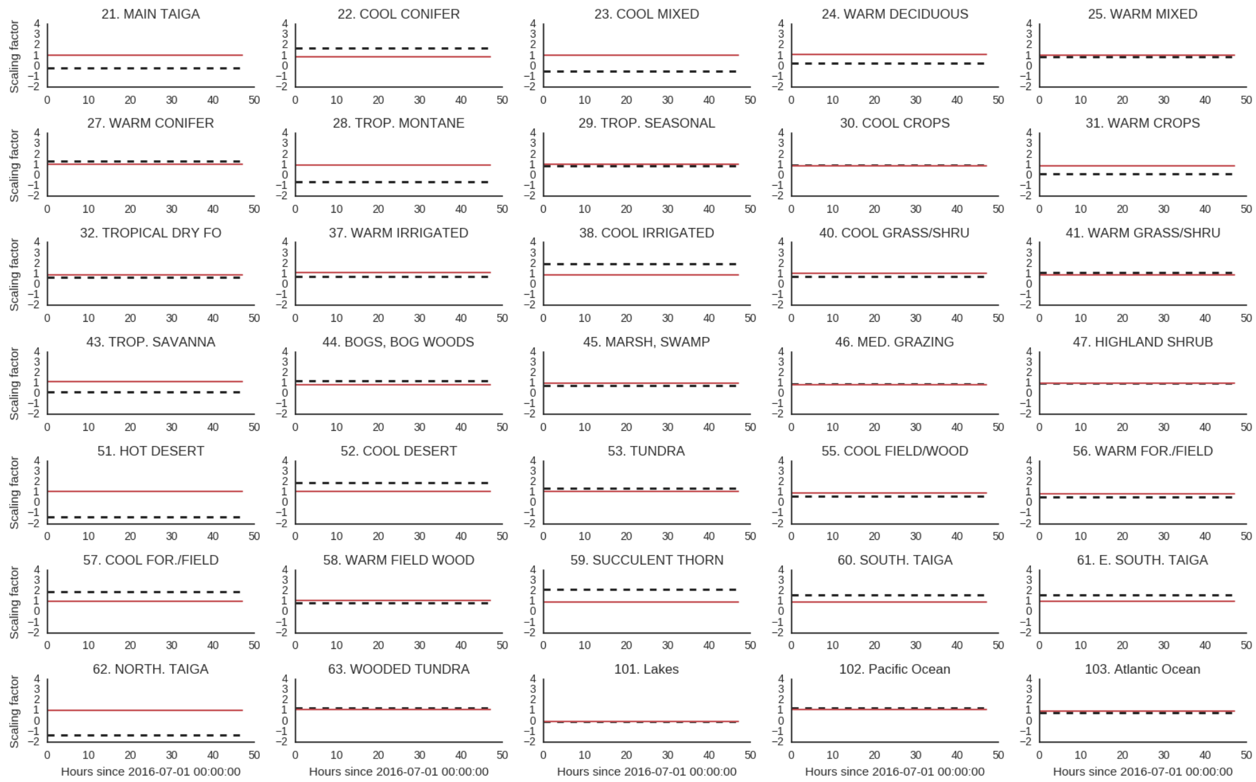
$$F_{true} = \alpha \cdot F_{prior}$$



Initial results from an idealized experiment

Dashed: Truth

Red: Initial guess

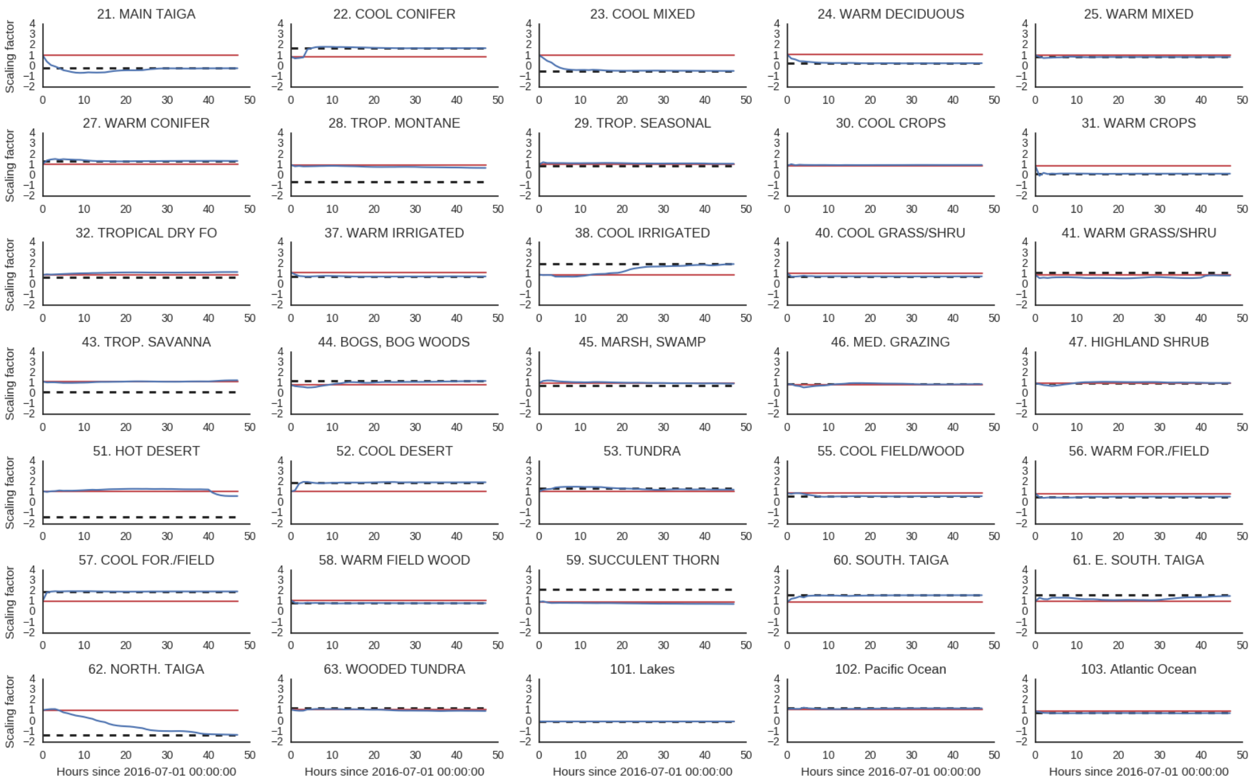


Initial results from an idealized experiment

Dashed: Truth

Red: Initial guess

Blue: Posterior



Initial results from an idealized experiment

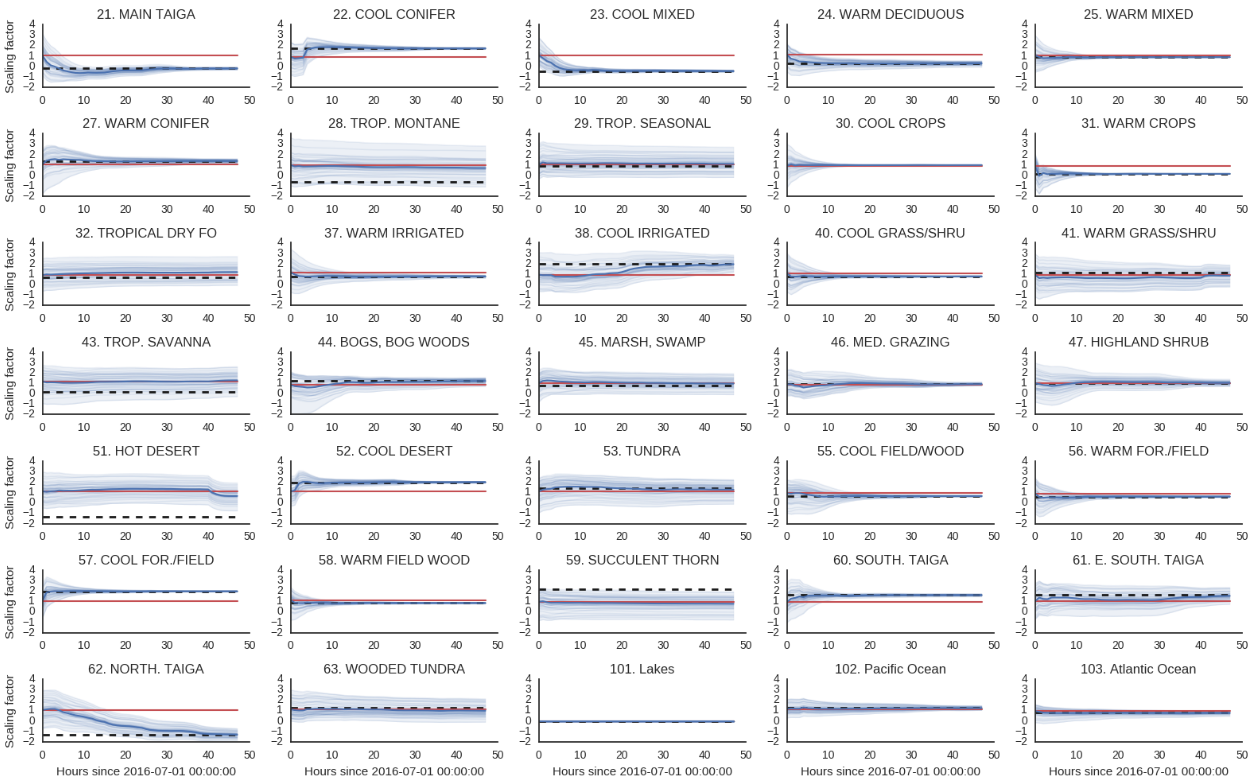
Dashed: Truth

Red: Initial guess

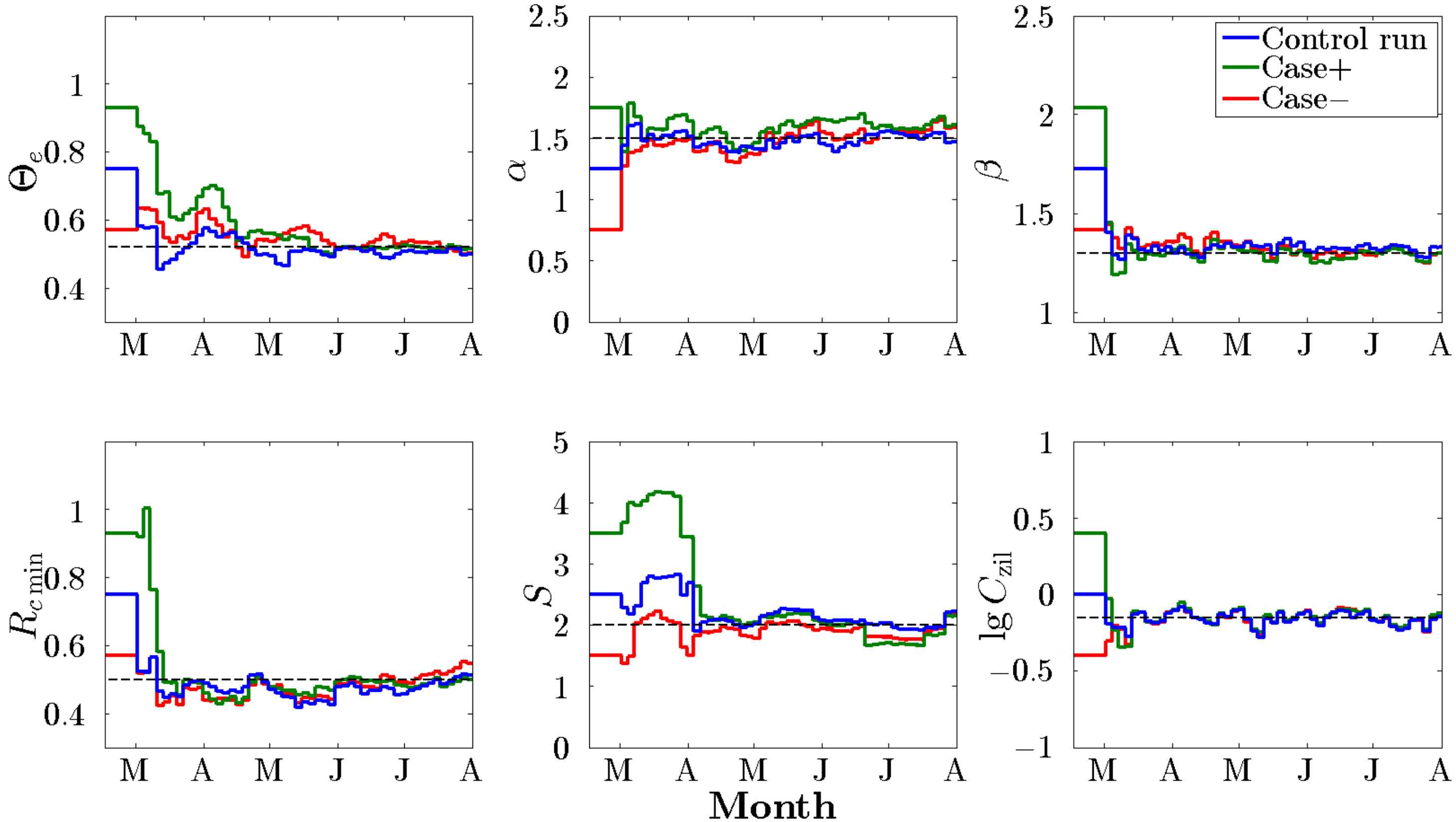
Blue: Posterior

Blue shading:

Ensemble members

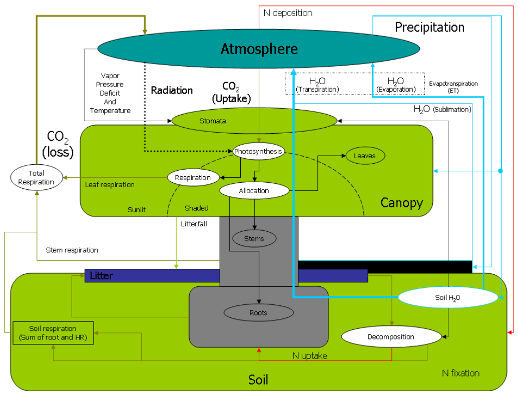


Towards Improved High-Resolution Land Surface Hydrologic Reanalysis Using a Physically-Based Land Surface Hydrologic Model and Data Assimilation



Shi et al. 2014a, Journal of Hydrometeology
Shi et al. 2014b, Water Resources Research
Shi et al. 2015, *Advances in Water Resources*

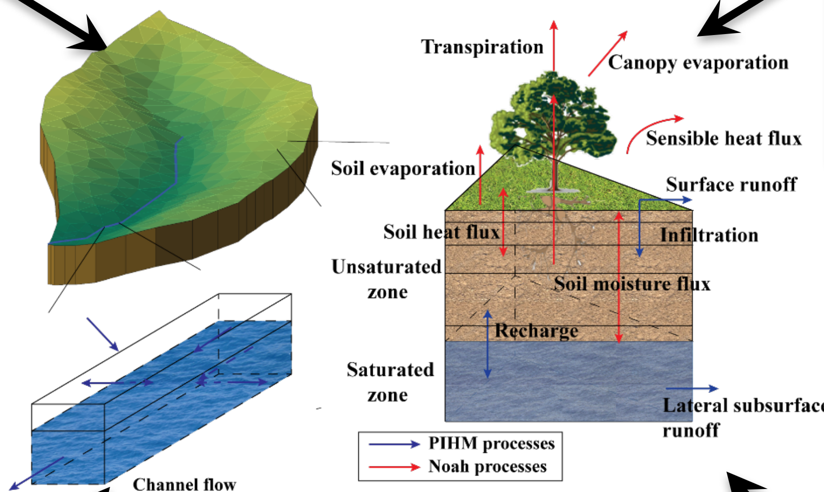
Coupled Hydro-Biogeochemistry Data Assimilation & Parameter Estimation

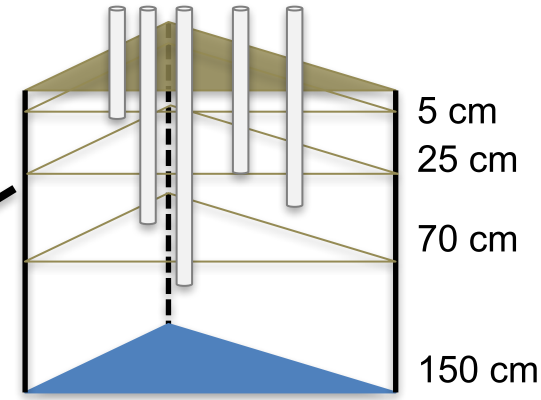


Forest Ecosystem Model
Biome-BGC

Crop Ecosystem Model Cycles

Flux-PIHM Data Assimilation System

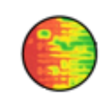




Geochemical Box Model
WITCH


Mineral dissolution / Precipitation


Adsorption Ion Exchange


Advection Diffusion/Dispersion

Reactive Transport Module

Courtesy of Yuning Shi

PENNSTATE

ESSPE: Ensemble-based Simultaneous State and Parameter Estimation

A generalized data assimilation software infrastructure for geoscience data-model integration

PSU PIs: F Zhang, L Li, S Brantley, W Brune, A Mejia, S Greybush & L Bao; NCAR PIs: J Anderson, D. Gochis & J. Richter

Numerical Models

Example earth processes for demo

Watershed biogeochemistry

local watershed model: Flux-PIHM

reactive transport model: RT-Flux-PIHM

Hydrology & land surface

regional community model: WRF-Hydro

Atmospheric chemistry

box chemistry model: RACM2

regional chemistry model: WRF-Chem

Earth and climate systems

community earth system model: CESM

Selected user contributed models

PSU EnDA ↔ **NCAR DART**

ESSPE

Statistical algorithms (EnKF, …)

Computing I/O interfaces Data protocols

Toolkits Modules User guide

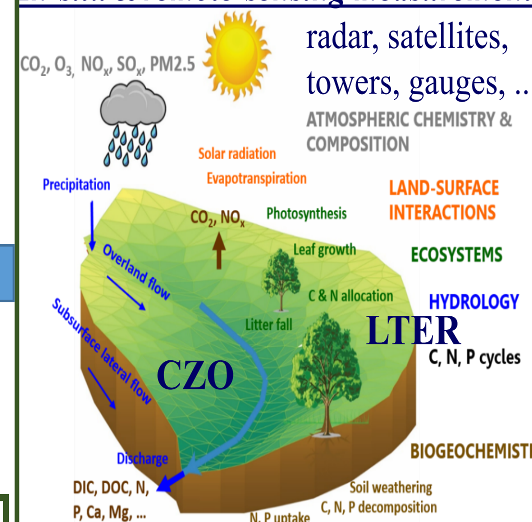
$$x^a_n = x^f_n + K\left(y^o - H_n x^f_n\right)$$

$$K = P^f H_n^T \left(H_n P^f H_n^T + R\right)^{-1}$$

$$x^f_n = \begin{bmatrix} x^f \\ \alpha^f \end{bmatrix} \begin{array}{l} \leftarrow \text{State variables} \\ \leftarrow \text{Model parameters} \end{array}$$

Observational Data

In-situ & remote-sensing measurements



GLEON USGS NEON

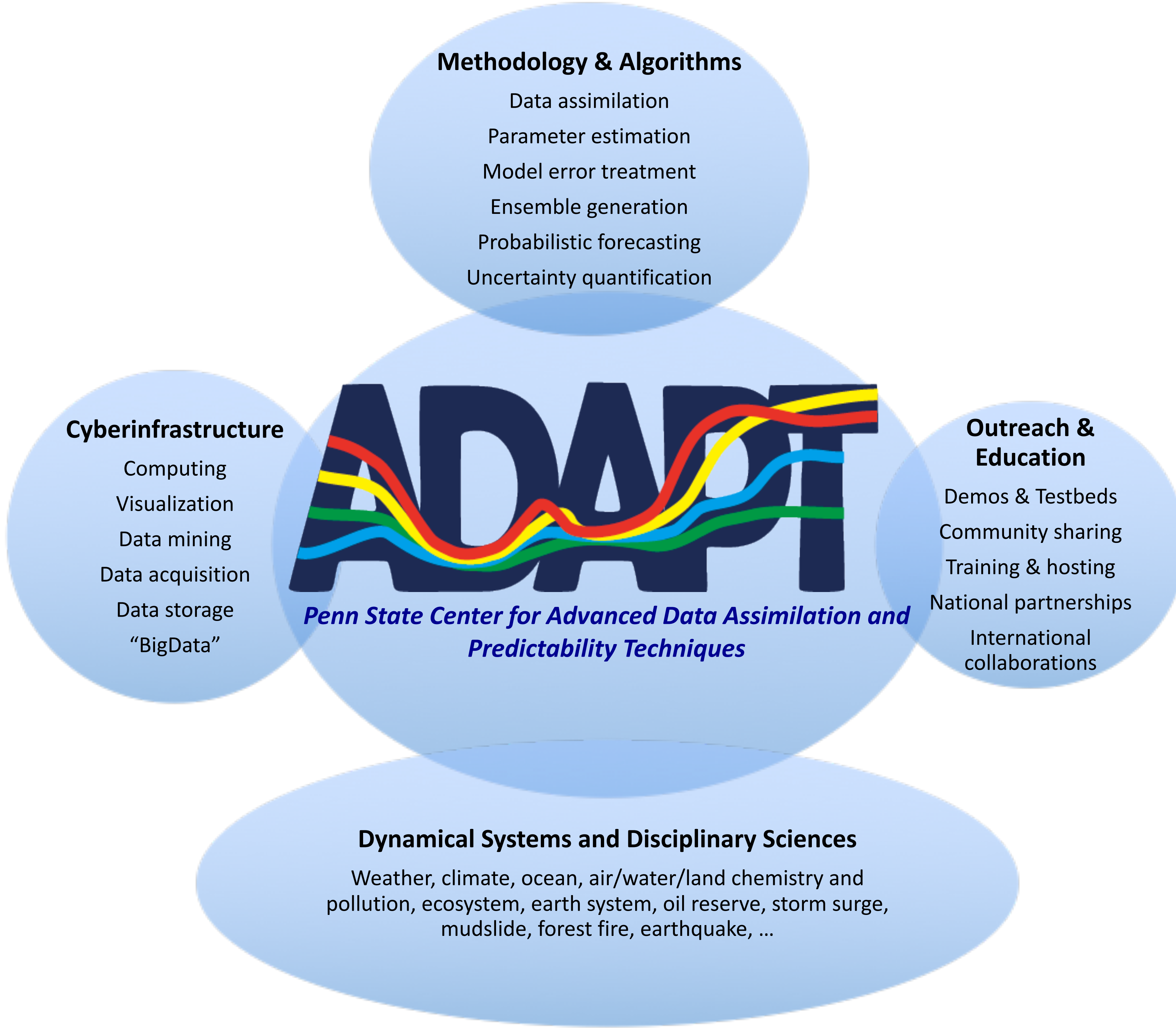
Data-Model Integration and Uncertainty Quantification

Estimating key parameters | Integrated 4-D data analysis

Identifying key processes | Observing network design

Key Expected Outcomes and Deliverables

- *Improving models and analyses*: better physics parameterizations, integrated analysis & understanding of geoscience processes
- *Better uncertainty quantification*: parameter sensitivity, observability & distinguishability; observing network optimization; predictability
- *Interconnect of geoscience community*: sharing data, model, algorithm & software; cross-validation of model & data across disciplines


Methodology & Algorithms
Data assimilation
Parameter estimation
Model error treatment
Ensemble generation
Probabilistic forecasting
Uncertainty quantification
Cyberinfrastructure
Computing
Visualization
Data mining
Data acquisition
Data storage
“BigData”
ADAPT
Penn State Center for Advanced Data Assimilation and Predictability Techniques
Outreach & Education
Demos & Testbeds
Community sharing
Training & hosting
National partnerships
International collaborations
Dynamical Systems and Disciplinary Sciences
Weather, climate, ocean, air/water/land chemistry and pollution, ecosystem, earth system, oil reserve, storm surge, mudslide, forest fire, earthquake, …

UCADA

University-NOAA Consortium on Advanced Data Assimilation

Partner Universities: Pennsylvania State University (PSU), University of Maryland (UMD), University of Oklahoma (OU) and University of Wisconsin (UW)

We propose to establish UCADA as a joint consortium between NOAA and universities seeks to integrate and enhance the existing strength and expertise in cutting-edge data assimilation (DA) research within and across operational and academic communities. UCADA will not only foster collaborations between NOAA scientists and university researchers, train the next-generation data assimilation specialists, but will also champion the two-way interactive intellectual exchanges in both research-to-operation (R2O) and operation-to-research (O2R). The new R2O and O2R interactive paradigms will facilitate rapid transition of new research development from the academic community to NOAA operations while the university researchers make concerted and direct efforts in seeking solutions to challenging DA issues emerged from operations.

UCADA Objectives

- Design and develop advanced and efficient DA algorithms for the next-generation operational NWP models from global to convective scales building on the strengths of the existing ensemble, adjoint and variational methods including various hybrids.

- Implement these advanced methods with fast, efficient numerical solvers and parallel computing capability under a community-consensus, object-oriented software framework that will be suitable for the current and next-generation NWP models. A potential candidate common DA software to be adopted is the one being developed at the Joint Center for Satellite Data Assimilation (JCSDA).

- Apply the advanced DA methods to assess the observability and predictability of various dynamical systems of interest, to improve the accuracy and design of various forecast systems, to assess the effectiveness and impact of the existing observing networks, to design the most cost-effective future observing systems through observing system simulation experiments and/or pilot real-time real-data predictions. A particular emphasis will be existing and forthcoming observations from satellites including the cloudy/rainy radiance data that are currently underutilized in operations.

- Serve as an intellectual hub for facilitating collaborative research between NOAA and universities, for attracting national and international visitors and scholars, for training and preparing graduate students and postdocs to be the next-generation data assimilation experts with strong ties to the operational communities, and as a testbed for rapidly transferring research to operations.