2008/11/3

Ensemble data assimilation at Japanese Met. Agency

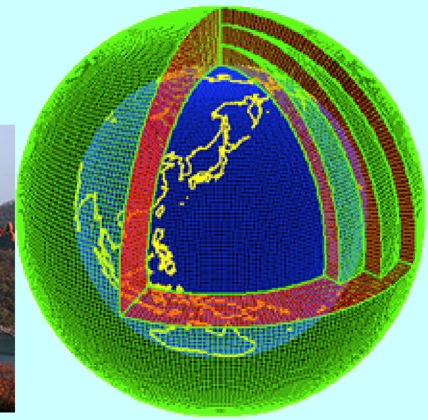
Takemasa Miyoshi
Numerical Prediction Division, JMA

Role of data assimilation

Major Methodologies in Atmos & Oceanic Sciences

Observing studies

Modeling studies



Data Assimilation

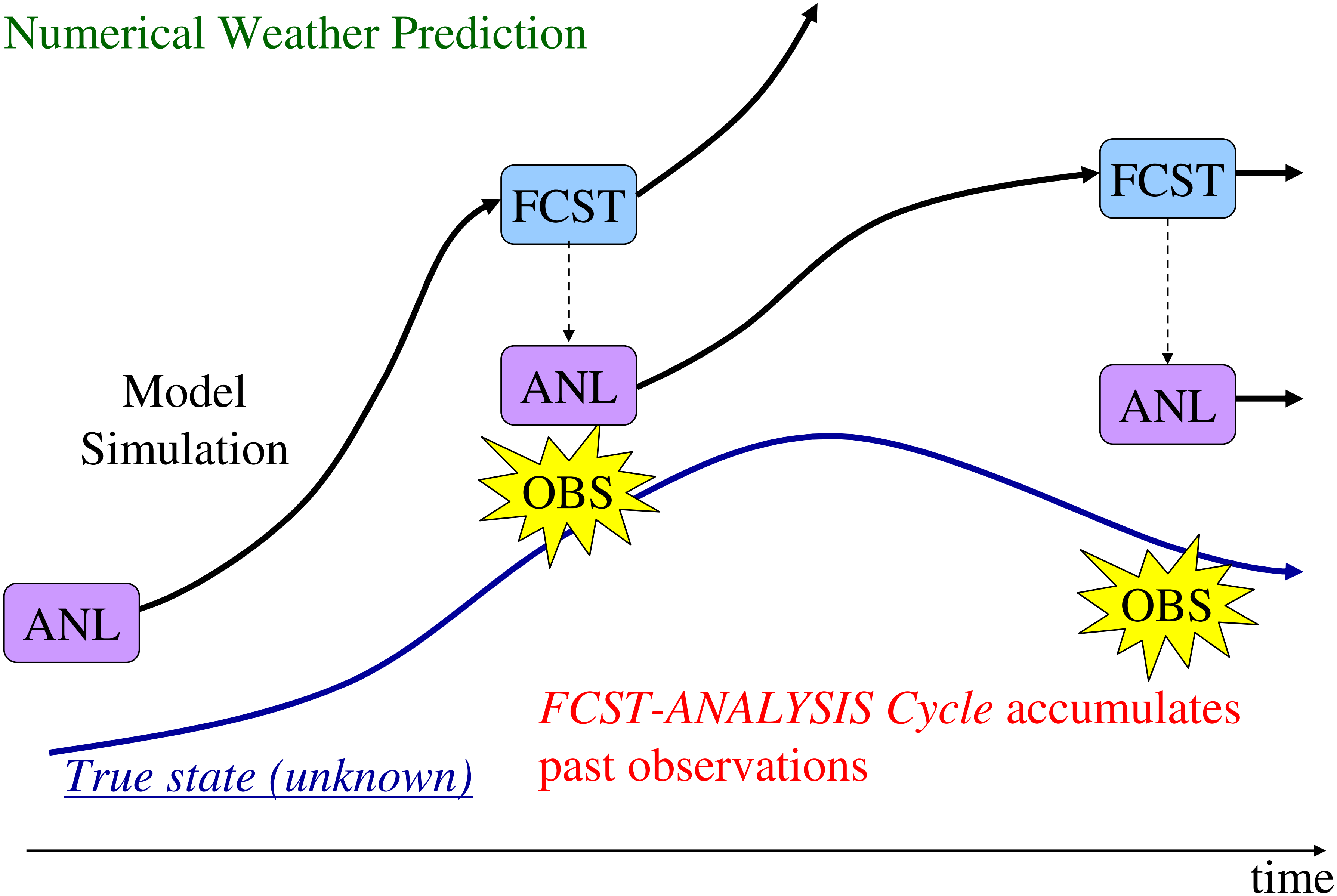
Initialize model predictions; essential in NWP

Observing scientists get dynamically consistent 4-D analysis

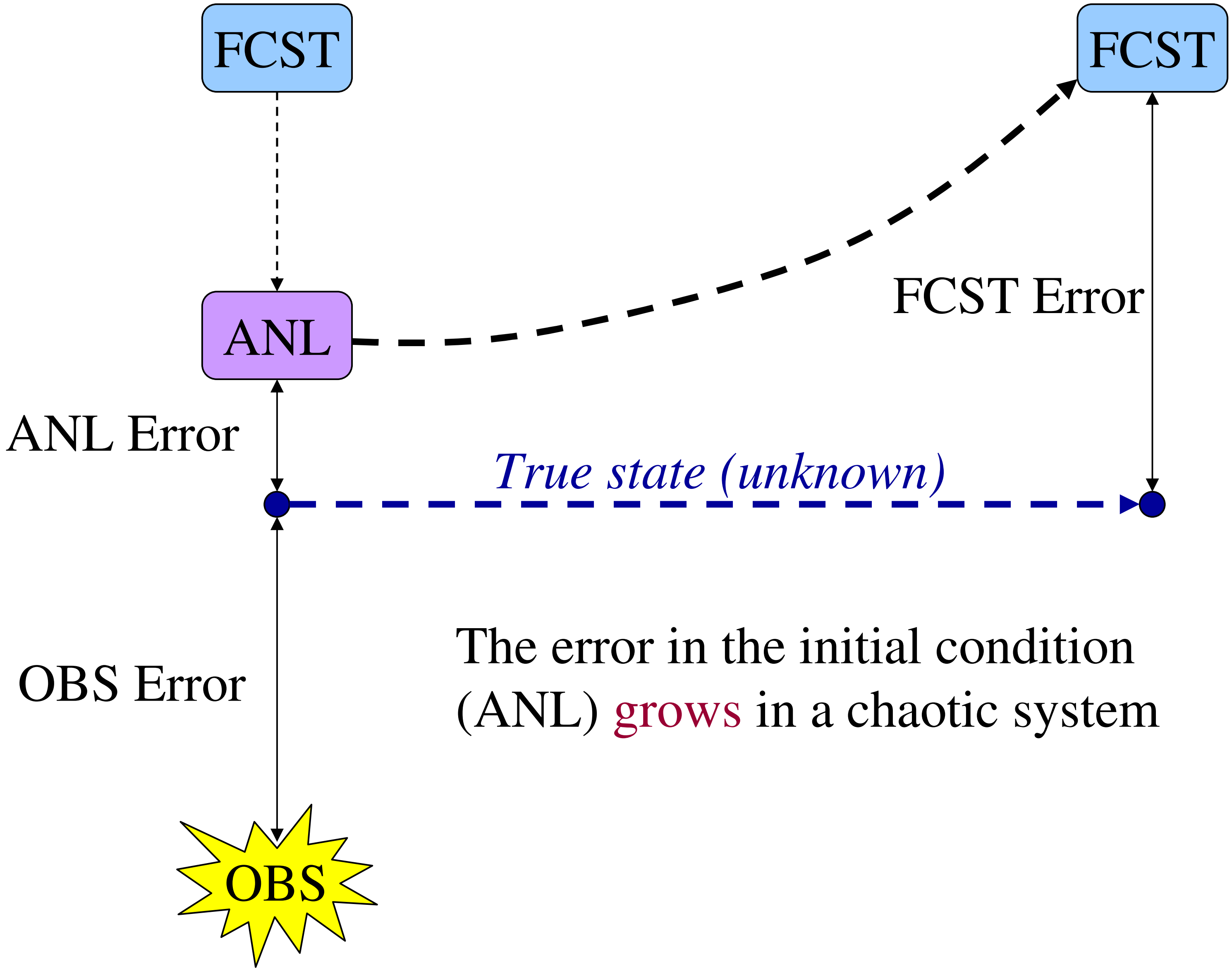
Modeling scientists get information of real nature

Synergy between two major fields helps scientific advancement

Data assimilation in NWP

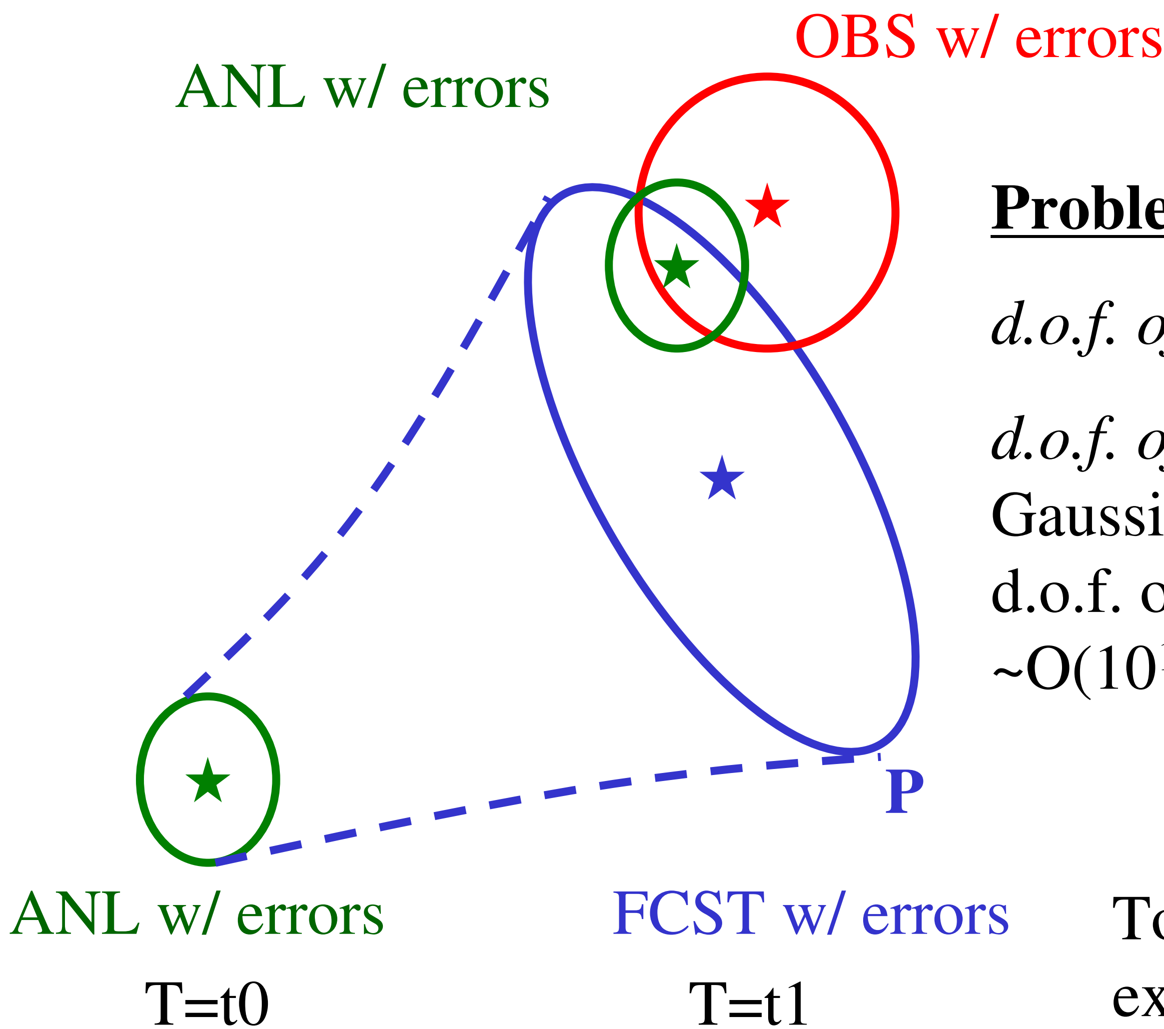


Data Assimilation = Analysis



The error in the initial condition (ANL) grows in a chaotic system

Probabilistic view



<u>Problem :</u>

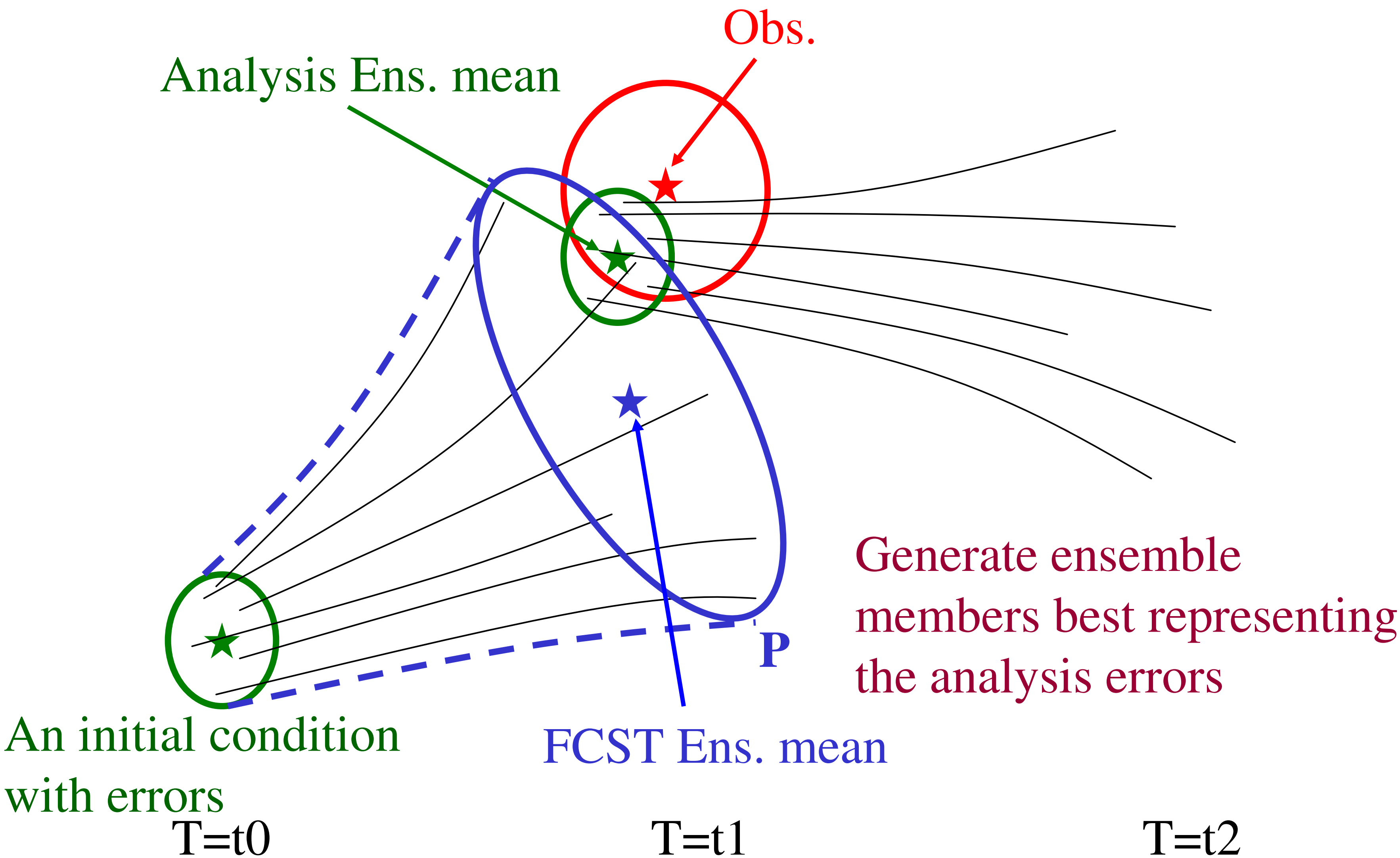
d.o.f. of the system: ~O(10^6)

d.o.f. of the error: even Gaussian distribution has d.o.f. of the covariance ~O(10^{12})

↓

Too large to express explicitly

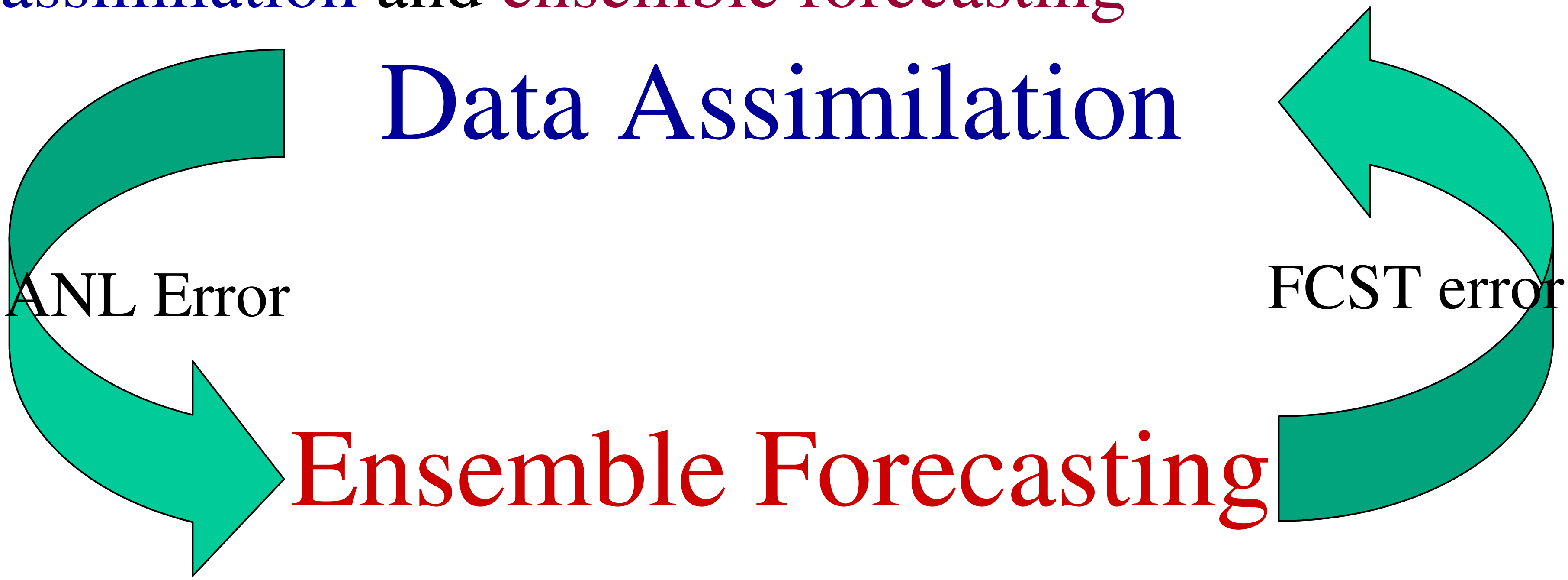
A schematic of EnKF



EnKF = ensemble fcst. + ensemble update

A core concept of EnKF

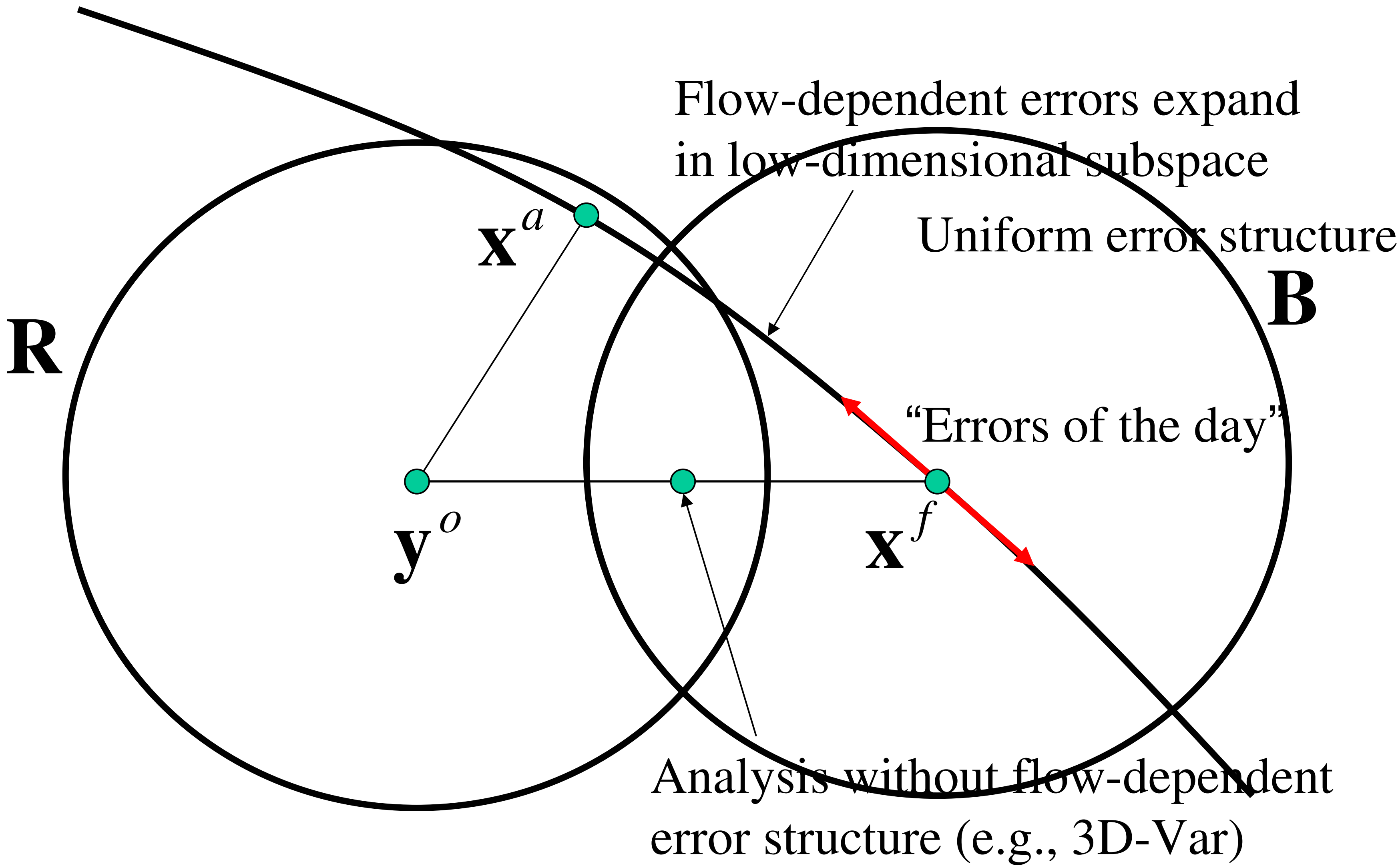
Complementary relationship between data assimilation and ensemble forecasting



This cycle process = EnKF

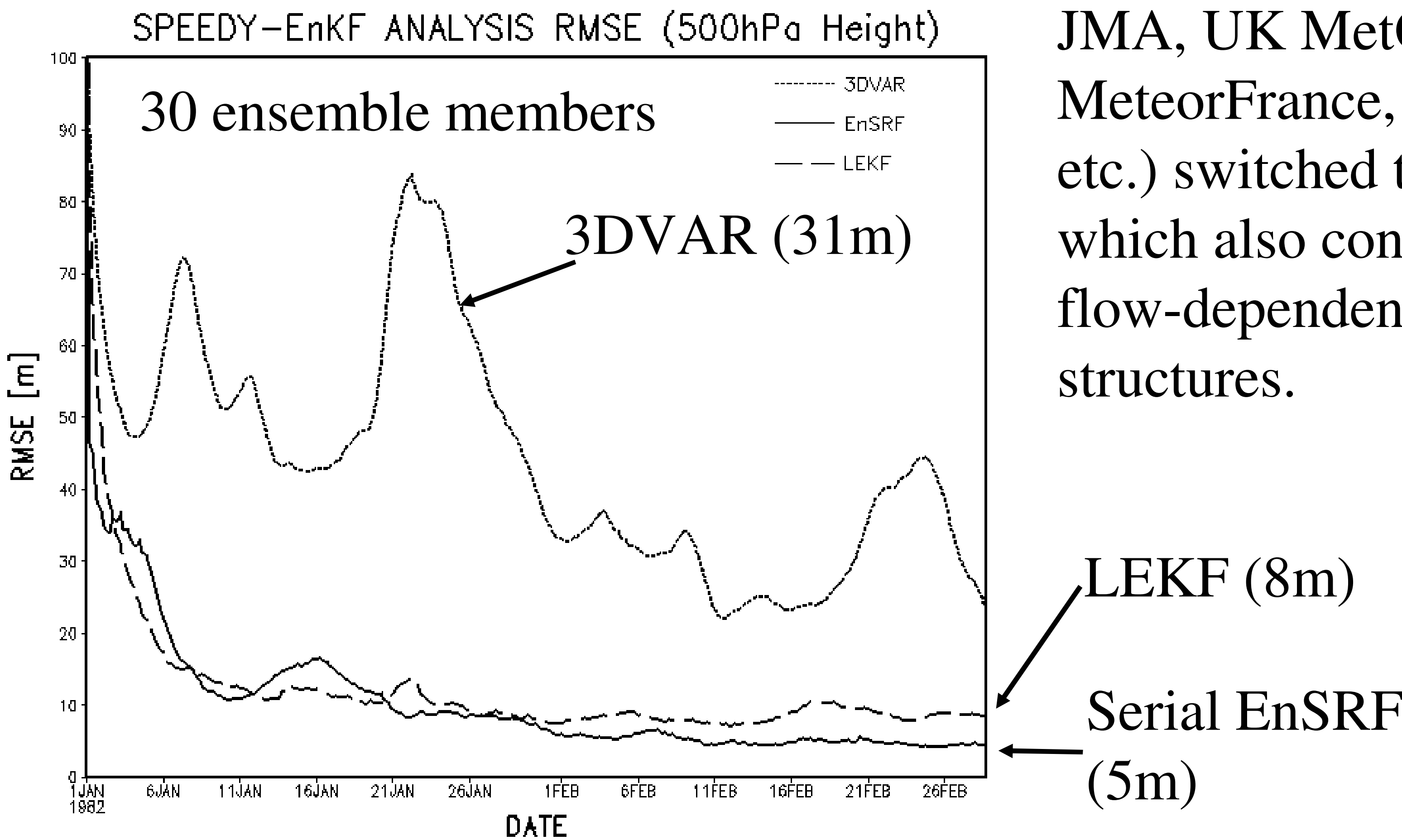
Analyze with the flow-dependent forecast error, ensemble forecast with initial ensemble reflecting the analysis error

Difference between EnKF and 3D-Var



An example of EnKF analysis accuracy

EnKF is advantageous to traditional data assimilation methods including 3D-Var, currently in operations at several NWP centers.



Many centers (ECMWF, JMA, UK MetOffice, MeteorFrance, Canada, etc.) switched to 4D-Var which also considers flow-dependent error structures.

EnKF - summary

- EnKF considers flow-dependent error structures, or the "errors of the day"
  - "advanced" data assimilation method
    - 4D-Var is also an "advanced" method. How different?

- EnKF analyzes the analysis errors in addition to analysis itself
  - "ideal" ensemble perturbations

EnKF vs. 4D-Var

	EnKF	4D-Var
"advanced" method?	Y	Y
Simple to code?	Y	N (e.g., Minimizer)
Adjoint model?	N	Y
Observation operator	Only forward (e.g., TC center)	Adjoint required
Asynchronous obs?	Y (4D-EnKF)	Y (intrinsic)
Analysis errors?	Y (ensemble ptb)	N
Limitation	ensemble size	Assim. window
	EnKF with infinite ensemble size and 4D-Var with infinite window are equivalent.	

LETKF (Hunt 2005; Hunt et al. 2007; Ott et al. 2004)

- Two categories of the EnKF (Ensemble Kalman Filter)

Perturbed observation (PO) method	Square root filter (SRF)
Classical	Relatively new
Already in operations (Canadian EPS)	Not in operations yet
Additional sampling errors by PO	No such additional sampling errors

- LETKF (Local Ensemble Transform Kalman Filter)
  - is a kind of ensemble square root filter (SRF)
  - is efficient with the parallel architecture

KF and EnKF

Kalman Filter

$\mathbf{x} : [N] \quad \mathbf{P} : [N \times N]$

Forecast equations

$$\mathbf{x}_i^f = M(\mathbf{x}_{i-1}^a)$$

$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{P}_{i-1}^a \mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q}$$

Kalman gain

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

Analysis equations

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

$$\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{P}_i^f$$

Ensemble Kalman Filter

$\mathbf{X} : [N \times m]$

Ensemble forecasts

$$\mathbf{X}_i^f = [M(\mathbf{x}_{i-1}^{a(1)}) | \cdots | M(\mathbf{x}_{i-1}^{a(m)})] \equiv M(\mathbf{X}_{i-1}^a)$$

Approximated by $\mathbf{P}_i^f \approx \dfrac{\delta\mathbf{X}_i^f (\delta\mathbf{X}_i^f)^T}{m-1}$

$$\delta\mathbf{Y} = H(\delta\mathbf{X}) : [p \times m]$$

$$\mathbf{K}_i = \delta\mathbf{X}_i^f (\delta\mathbf{Y})^T [\delta\mathbf{Y}(\delta\mathbf{Y})^T + (m-1)\mathbf{R}]^{-1}$$

[pxp] matrix inverse

Solve for the ensemble mean

Ensemble perturbations

$$\delta\mathbf{X}^a = [(m-1)\mathbf{P}^a]^{1/2}$$

LETKF algorithm (Hunt, 2005, et al., 2007)

$$\mathbf{P}^f \approx \frac{\delta\mathbf{X}^f(\delta\mathbf{X}^f)^T}{m-1} = \delta\mathbf{X}^f \underline{\tilde{\mathbf{P}}^f} (\delta\mathbf{X}^f)^T \qquad \tilde{\mathbf{P}}^f = \frac{\mathbf{I}}{m-1} : [m \times m]$$

In the space spanned by $\delta\mathbf{X}^f$

$$\tilde{\mathbf{P}}^a = [\underline{(m-1)\mathbf{I}/\rho + (\delta\mathbf{Y})^T\mathbf{R}^{-1}\delta\mathbf{Y}}]^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T$$

→ Eigenvalue decomposition: $\mathbf{U}\mathbf{D}\mathbf{U}^T : [m \times m]$

Analysis equations

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \delta\mathbf{X}^f\tilde{\mathbf{P}}^a(\delta\mathbf{Y})^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{H(\mathbf{x}^f)})$$

$$\delta\mathbf{X}^a = \delta\mathbf{X}^f[(m-1)\tilde{\mathbf{P}}^a]^{1/2} = \delta\mathbf{X}^f\sqrt{m-1}\,\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T$$

LETKF analysis

$$\mathbf{X}^a = \overline{\mathbf{x}}^f + \underline{\delta\mathbf{X}^f\left(\tilde{\mathbf{P}}^a(\delta\mathbf{Y})^T\mathbf{R}^{-1}(\mathbf{y}^o - \overline{H(\mathbf{x}^f)}) + \sqrt{m-1}\,\mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^T\right)}$$

Ensemble analysis increments

Research activities

MY MISSION:
Developing a next-generation data assimilation system to improve operational NWP at JMA

Path to operations

1. Develop and test LETKF (Hunt 2005; Hunt et al. 2007; Ott et al. 2004; Hunt et al. 2004) with the Earth Simulator
2. Develop LETKF with the JMA nonhydrostatic model
3. Develop LETKF with the JMA global model
4. Assess LETKF under the quasi-operational setup

Researches using the Earth Simulator LETKF system

- Experimental 1.5-yr reanalysis (ALERA)
- Collaborative work with observing scientists
- LETKF with the Earth Simulator coupled atmos-ocean GCM

Outline

- Developments of LETKF toward operations
  - Recent improvements
  - Quasi-operational comparison with 4D-Var
  - Probabilistic forecast skills

- Research projects with the Earth Simulator
  - Experimental Ensemble Reanalysis: ALERA
  - Collaboration with observing scientists

Developments toward operations

LETKF developments at JMA

- LETKF (Local Ensemble Transform Kalman Filter, U of MD, Hunt et al. 2007; Ott et al. 2004) has been applied to 3 models
  - AFES (AGCM for the Earth Simulator)

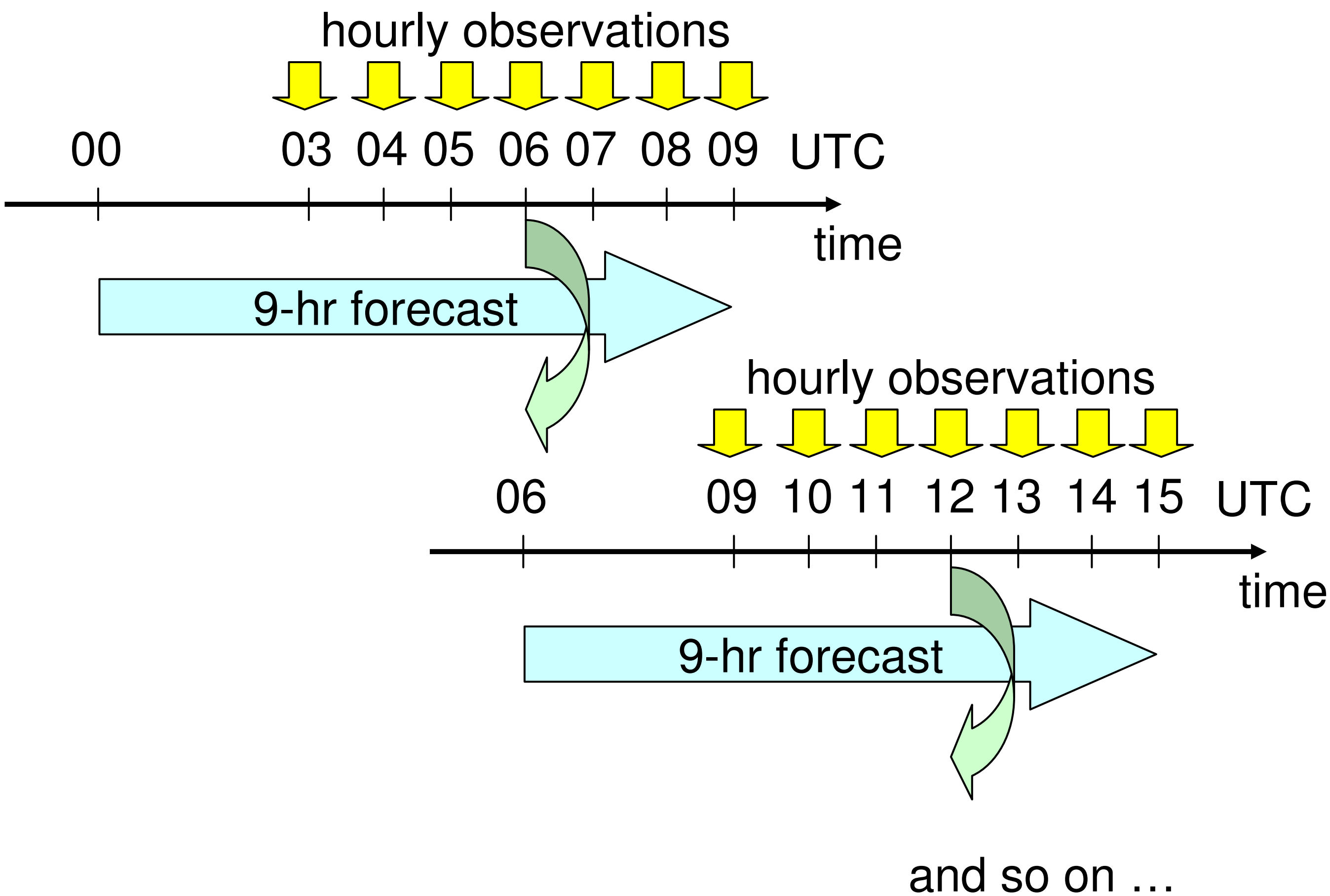
    Miyoshi and Yamane, 2007: *Mon. Wea. Rev.*, 3841-3861.

    Miyoshi, Yamane, and Enomoto, 2007: *SOLA*, 45-48.
  - NHM (JMA nonhydrostatic model)

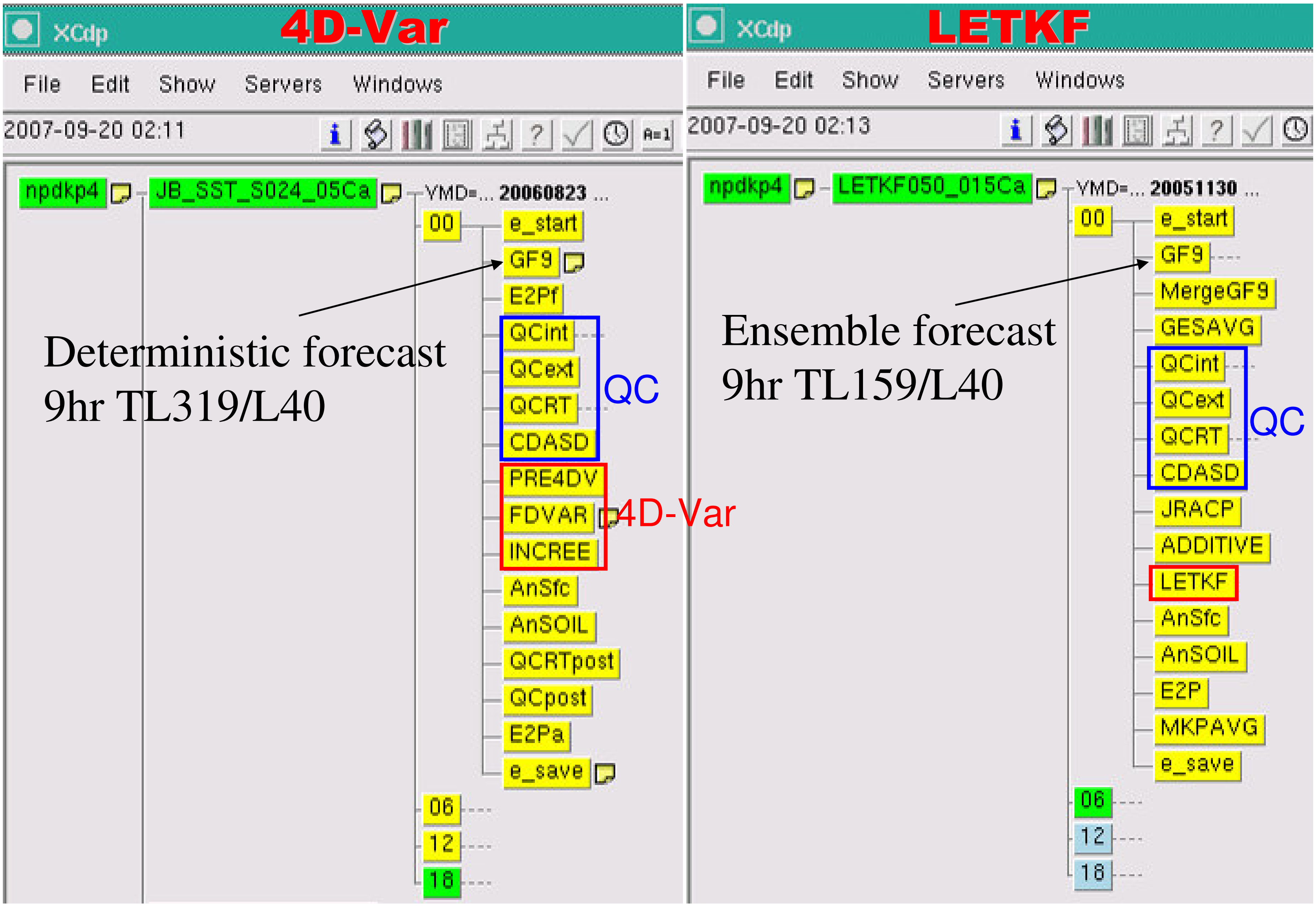
    Miyoshi and Aranami, 2006: *SOLA*, 128-131.
  - GSM (JMA global spectral model)

    Miyoshi and Sato, 2007: *SOLA*, 37-40.

Analysis-Forecast cycle experiment



Quasi-operational Experimental System



Recent improvements

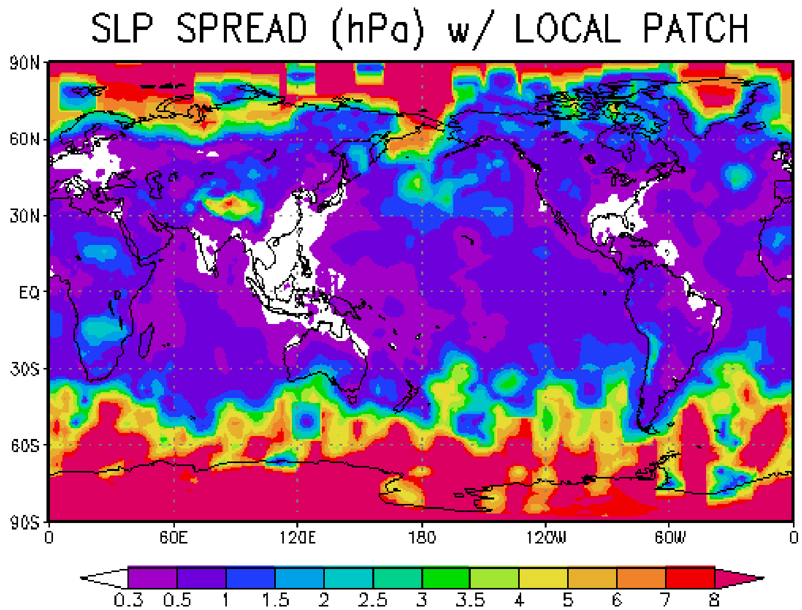
- Assimilation of satellite radiances
  - greatly improves the analysis accuracy

    Miyoshi and Sato, 2007: *SOLA*, 37-40.
- Removing local patches
  - solves the discontinuity problem near the Poles

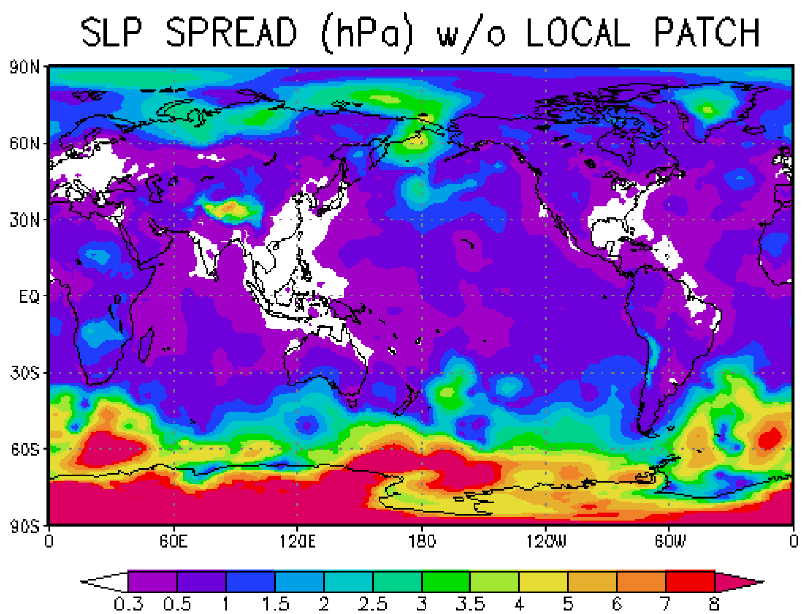
    Miyoshi et al., 2007: *SOLA*, 89-92.
- Efficient MPI parallel implementation
  - solves the load imbalance problem
  - accelerates by a factor of 3

    about 30% faster than operational 4D-Var with similar settings
- Adaptive satellite bias correction
  - a new idea analogous to the variational bias correction
  - showing great positive impact

LETKF without local patches



SLP analysis ensemble spread after the first analysis step

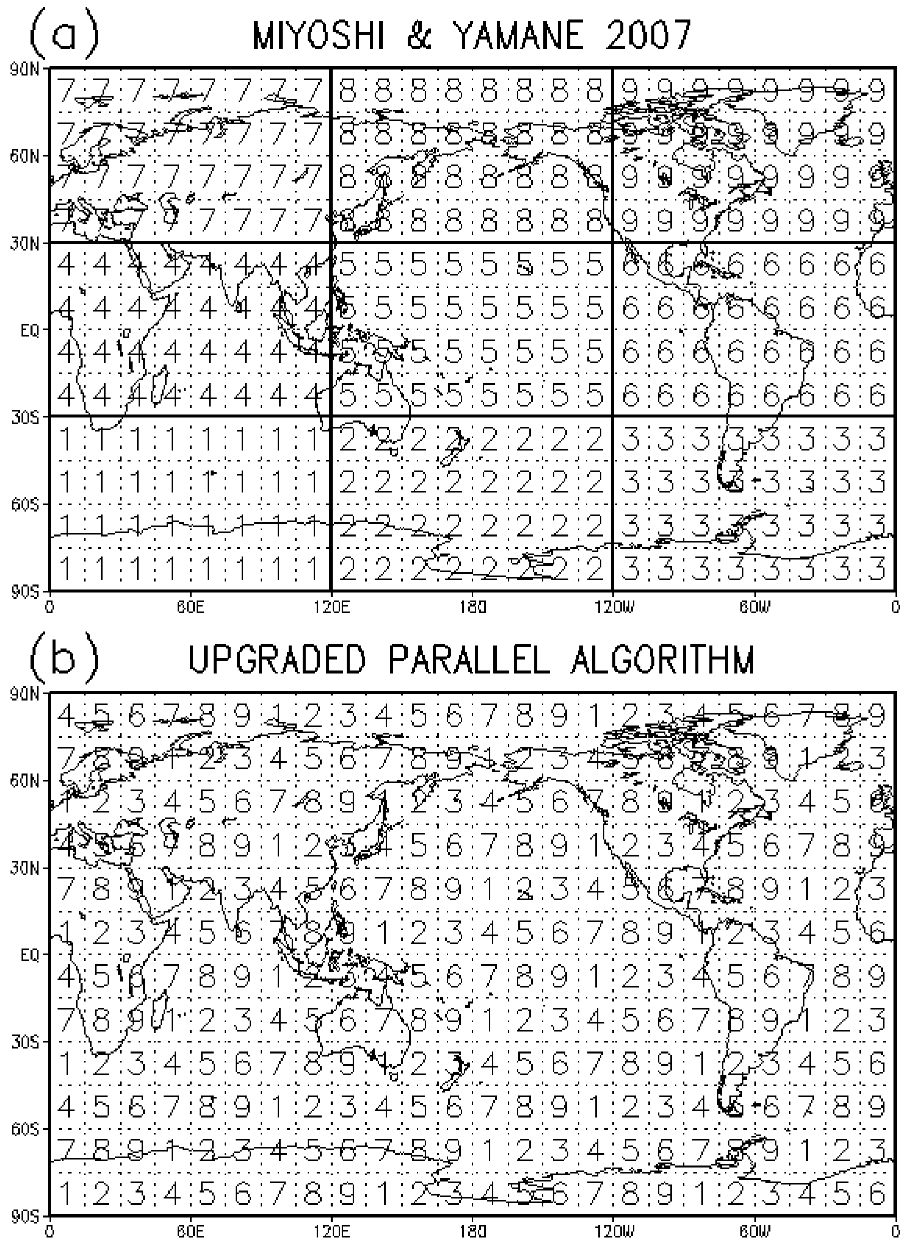


The discontinuities caused by the local patches disappear.

EARTH SIMULATOR

Miyoshi et al. 2007

Efficient parallel implementation

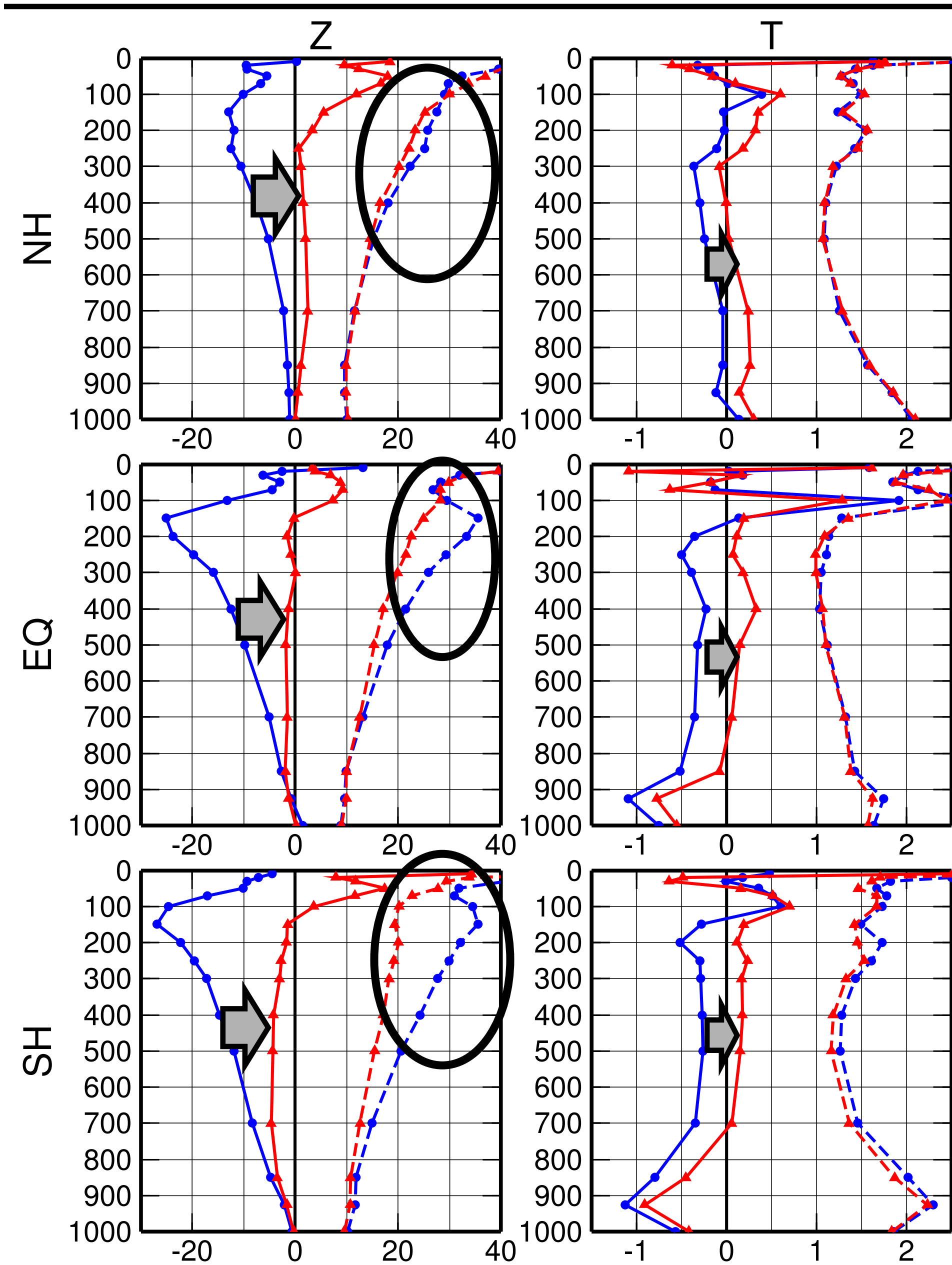


In the case of 9 comp. nodes

Irregular observing network causes significant load imbalances

Revising the node separation, we solved the load-imbalance problem almost completely; ~3 times faster computation

Impact by satellite radiances



RMSE and bias against radiosondes

Blue: w/o satellite radiances

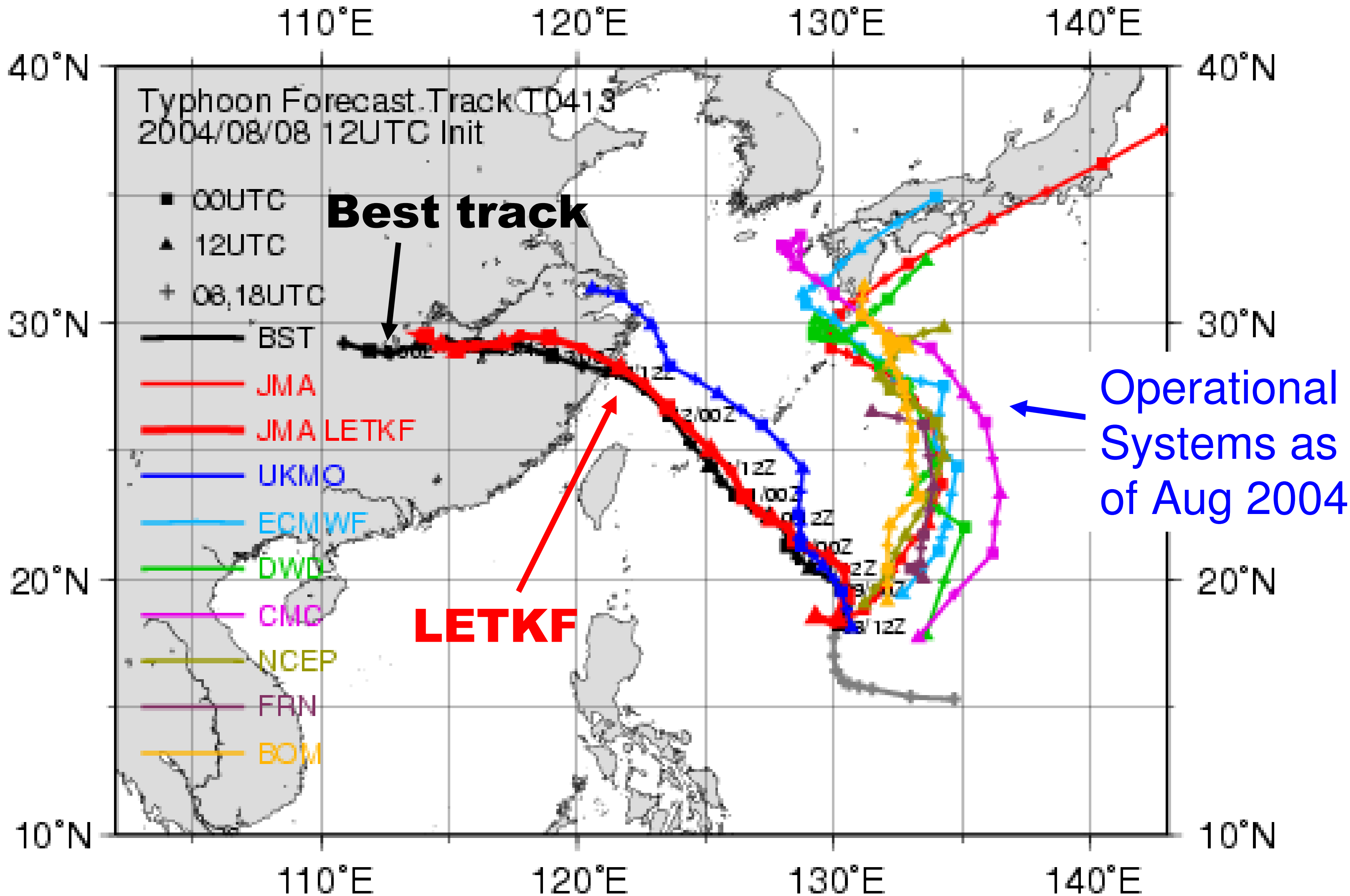
Red: w/ satellite radiances

Reduced negative bias of Z and T

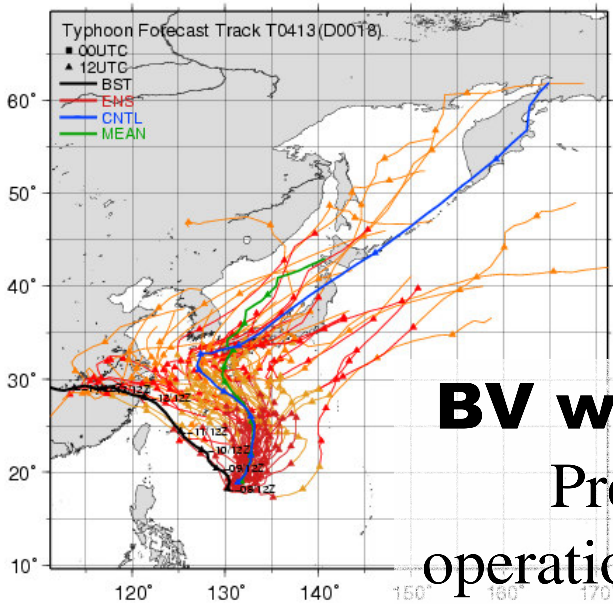
Reduced RMSE of Z in mid-upper troposphere (500-100hPa), especially in the SH and Tropics

20 members

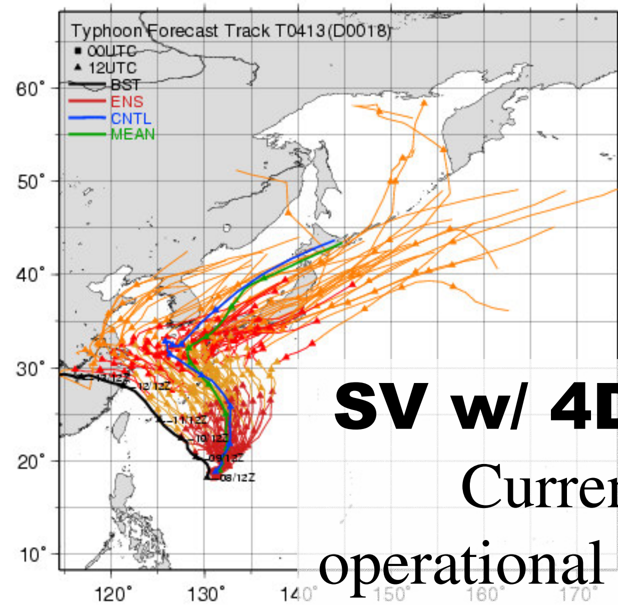
Typhoon Rananim, August 2004



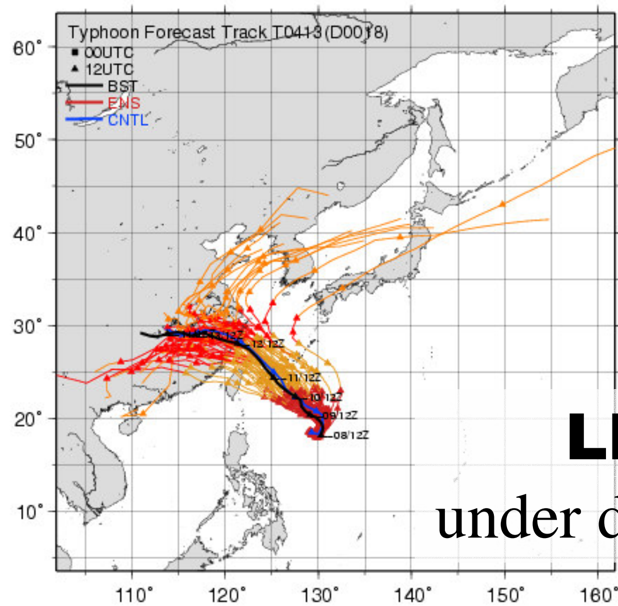
TC track ensemble prediction



BV w/ 4D-Var
Previous operational system



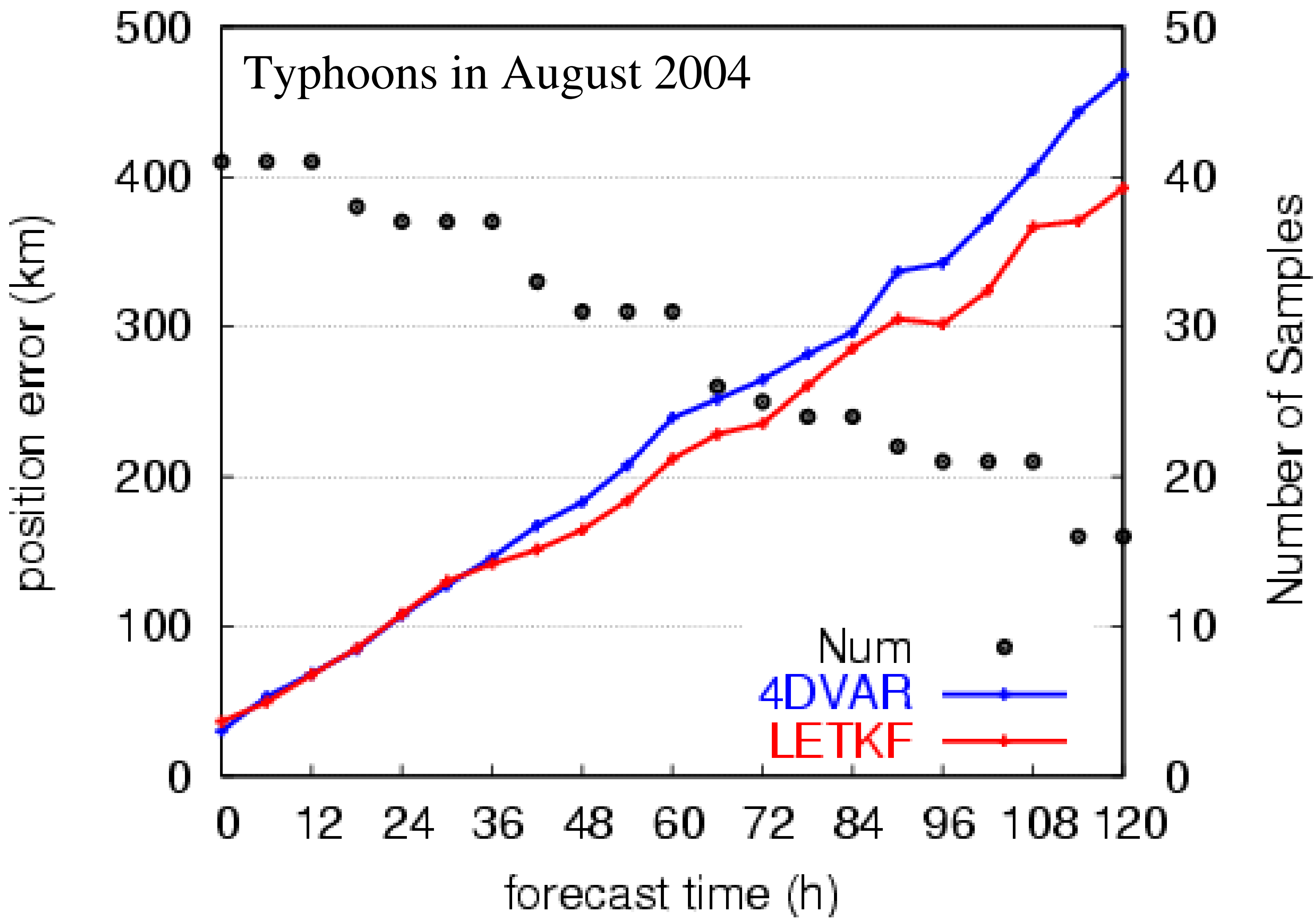
SV w/ 4D-Var
Current operational system



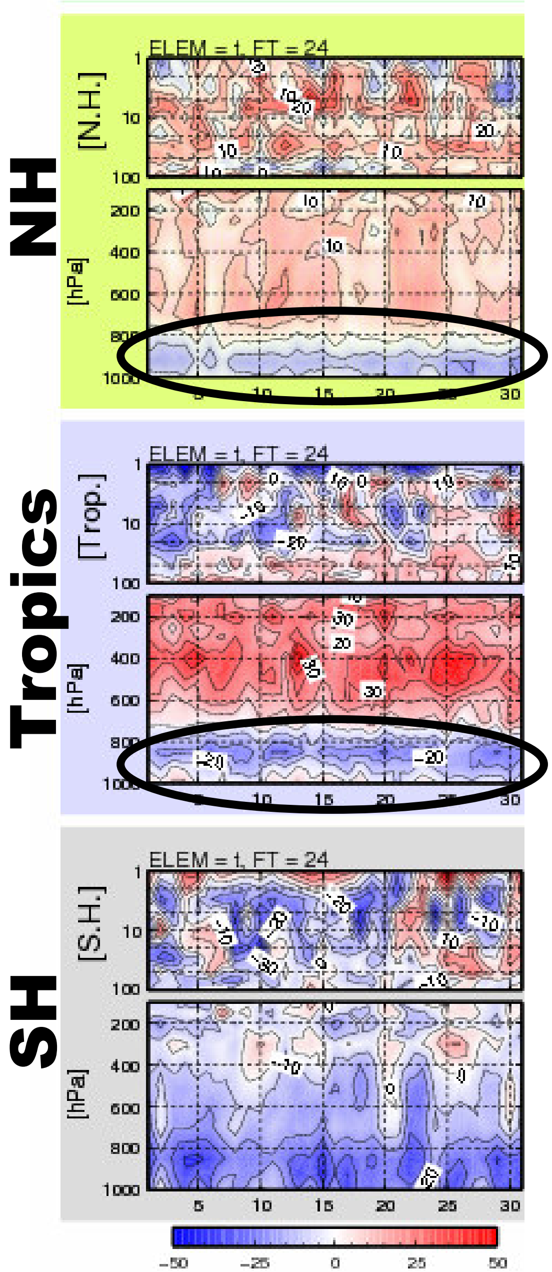
LETKF
under development

LETKF performs excellent in this typhoon case.

Statistical typhoon track errors



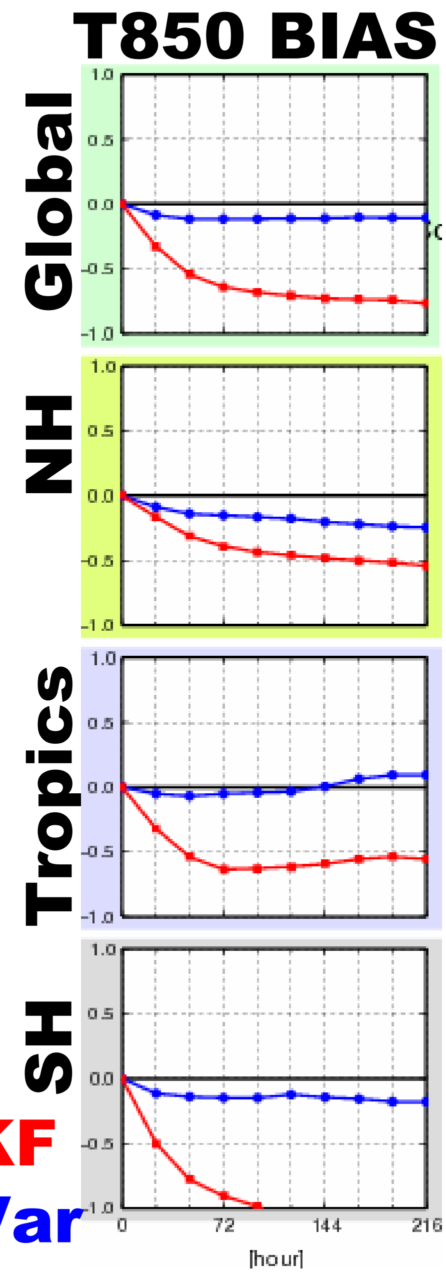
However, there was a problem



RED: LETKF is better
BLUE: 4D-Var is better

Period: August 2004

Significant low temperature bias in the lower troposphere



Red: LETKF
Blue: 4D-Var

Satellite radiance bias correction

Observation y has a bias b

$$b = b^{scan} + b^{air}$$

Air mass bias (dependent on atmospheric state)

Scan bias (constant)

Statistically estimated offline

Coefficients β of predictors p are estimated statistically

$$b^{air} = p^T \beta$$

- Zenith angle
- Surface temperature
- Constant
- etc.

Adaptive bias correction

Coefficients would change partly due to the deterioration of sensors

Allow temporal variation of the coefficients using data assimilation

Variational bias correction (e.g., Dee 2003; Sato 2007)

$$J(x) = \tfrac{1}{2}(x - x^f)B^{-1}(x - x^f)^T + \tfrac{1}{2}(y - Hx^f)R^{-1}(y - Hx^f)^T$$

$$J(x,\beta) = \tfrac{1}{2}(x - x^f)B^{-1}(x - x^f)^T + \tfrac{1}{2}\underline{(\beta - \beta^f)B_\beta{}^{-1}(\beta - \beta^f)^T}$$

$$+\tfrac{1}{2}(y \underline{- p^T\beta} - Hx^f)R^{-1}(y \underline{- p^T\beta} - Hx^f)^T$$

Find the minimizer β of the cost function J through the variational procedure

Adaptive bias correction with LETKF

Analytical solution of the variational problem: minimizer (x, β)

$$\begin{cases} \delta x = (B_x^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} (d - p^T \delta\beta) \\ \delta\beta = (B_\beta^{-1} + p R^{-1} p^T)^{-1} p R^{-1} (d - H\delta x) \end{cases}$$

Adaptive bias correction with LETKF

1. Solve the LETKF data assimilation problem first

$$\delta x = B_x H^T (H B_x H^T + R)^{-1} d = (B_x^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} d$$

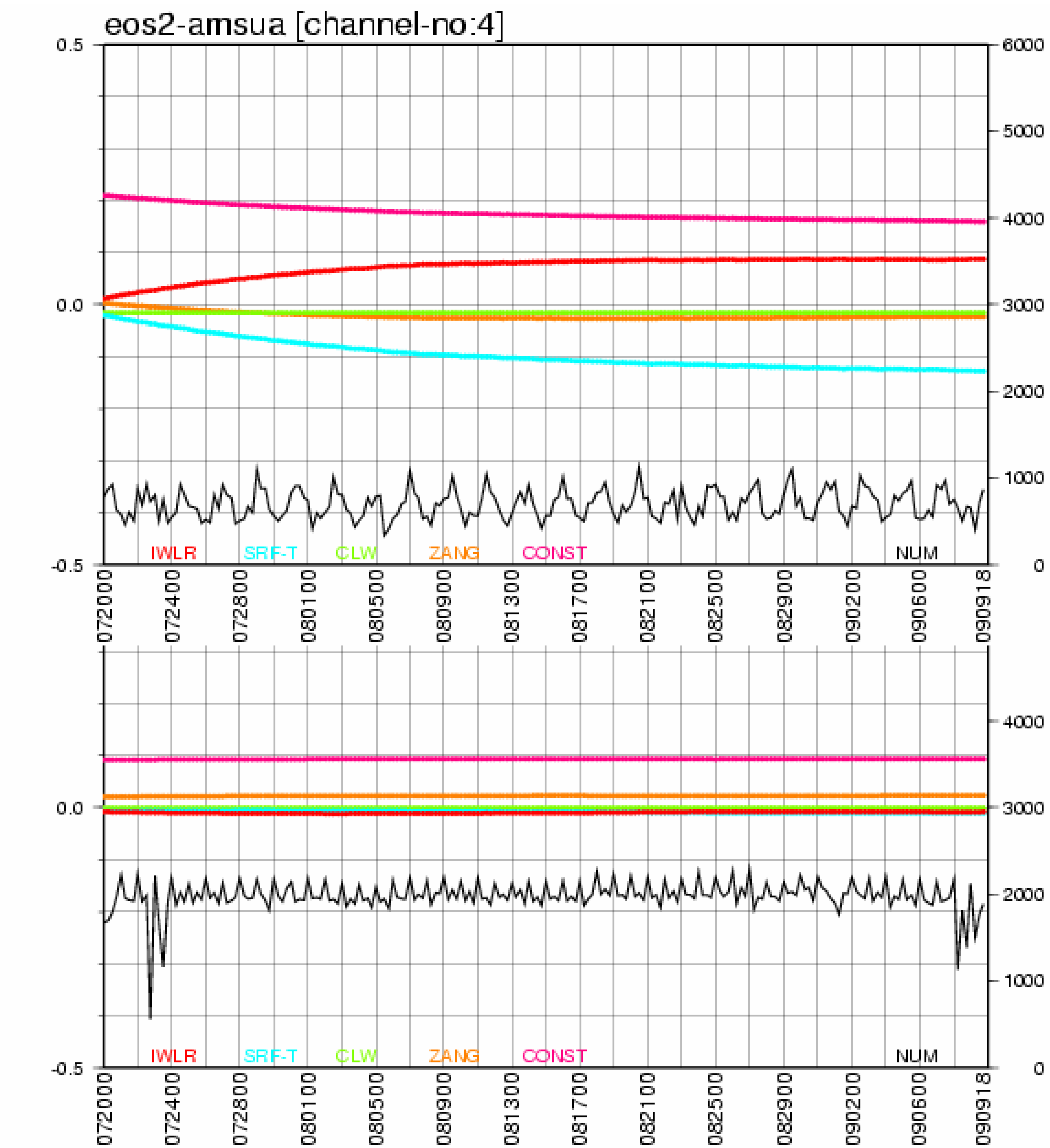
$- p^T \delta\beta$ difference

2. Solve the equation for β explicitly

$$\delta\beta = (B_\beta^{-1} + p R^{-1} p^T)^{-1} p R^{-1} (d - H\delta x)$$

This coincides with the variational BC

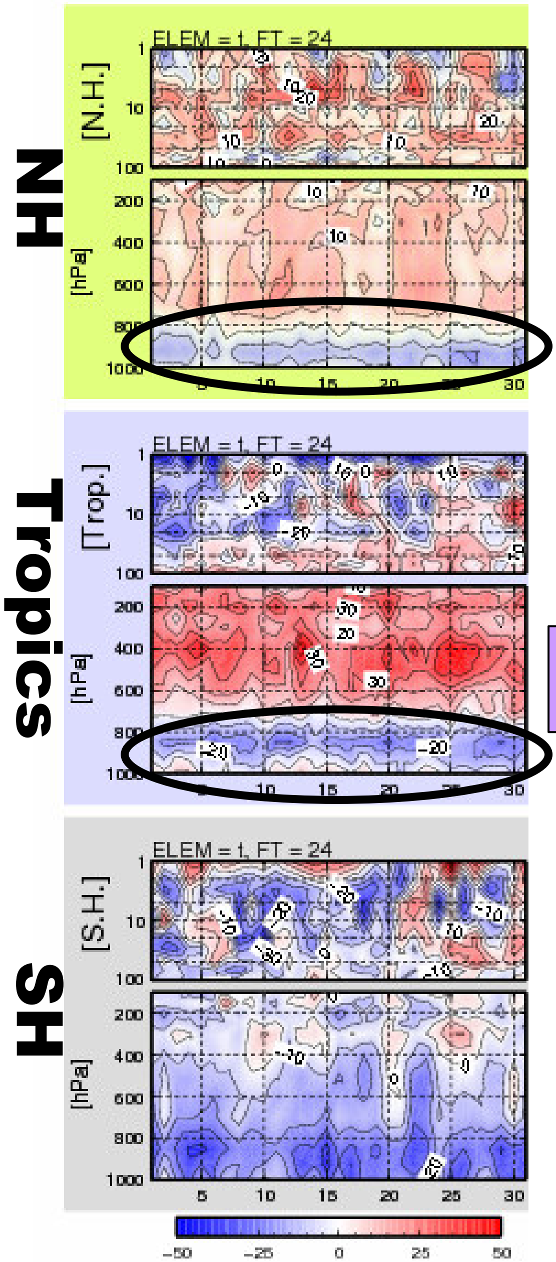
Time series of bias coefficients



AMSU-A 4ch (sensitive to middle-lower troposphere) indicates significant drift from those estimated by 4D-Var

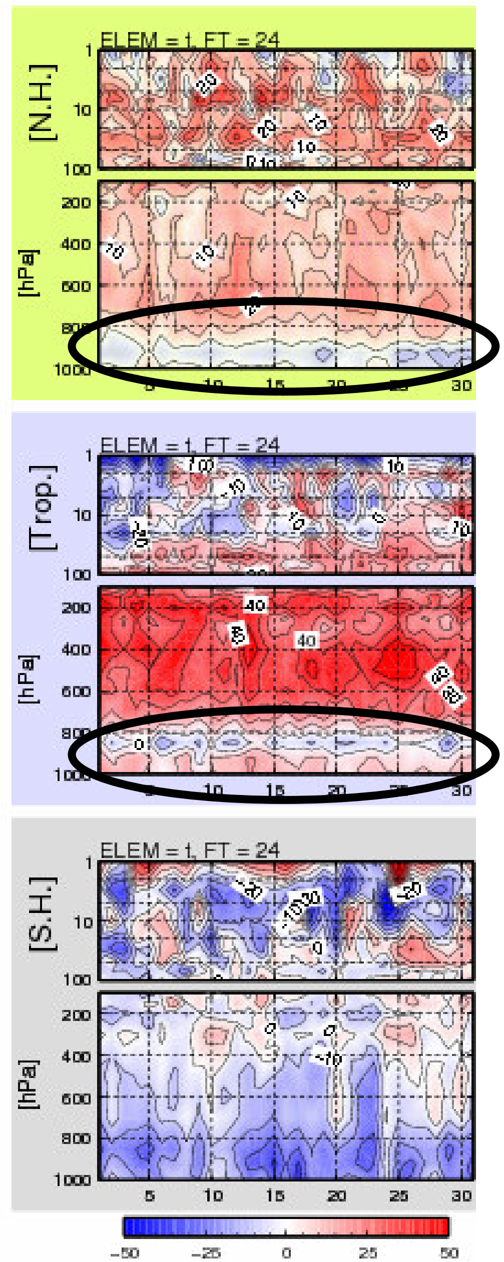
AMSU-A 6ch (sensitive to upper troposhere) and other sensors/channels indicate no significant drift

Impact by adaptive bias correction

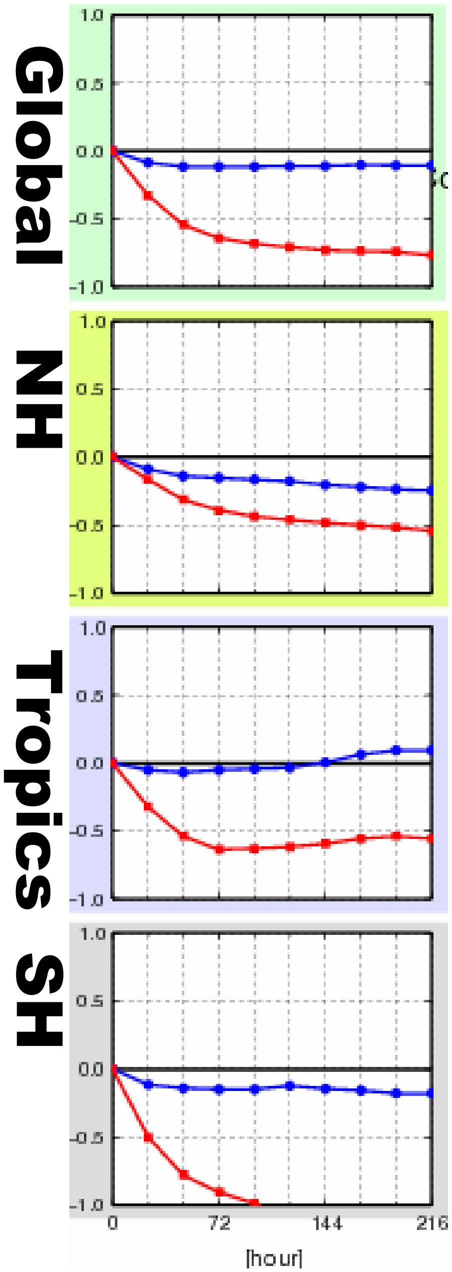


24hr temperature forecast error improvements relative to 4D-Var

RED: LETKF is better
BLUE: 4D-Var is better



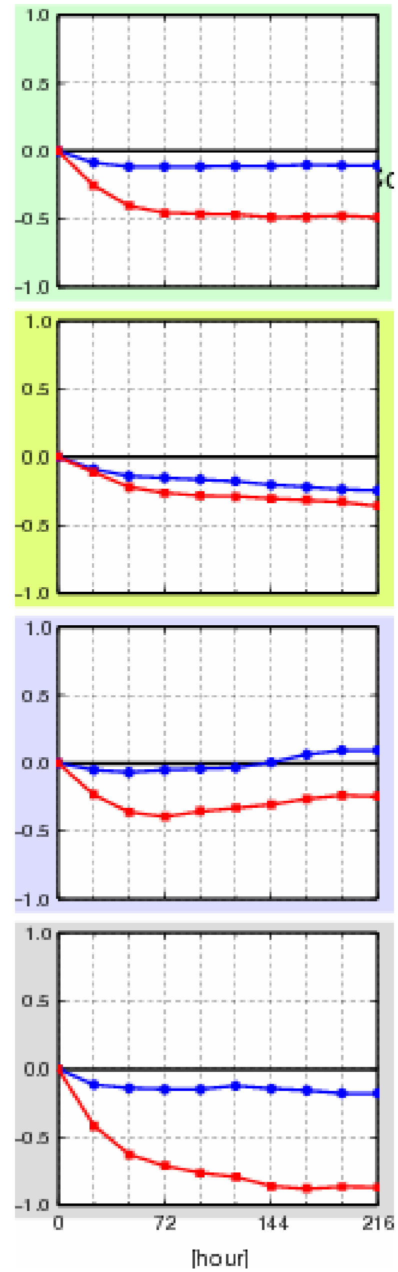
Bias reduction



T850 forecast bias
against initial condition

Red: LETKF
Blue: 4D-Var

Apply Adaptive BC



The improvements would be due to the bias reduction

Improvement (%) relative to 4D-Var

	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	–9.00	–10.45	–10.64	2.38	0.13
N. Hem.	–4.47	–2.95	–1.72	3.74	0.66
Tropics	0.48	–11.66	–17.60	11.69	9.88
S. Hem.	–10.90	–14.51	–13.00	–1.52	–3.81

Apply adaptive bias correction

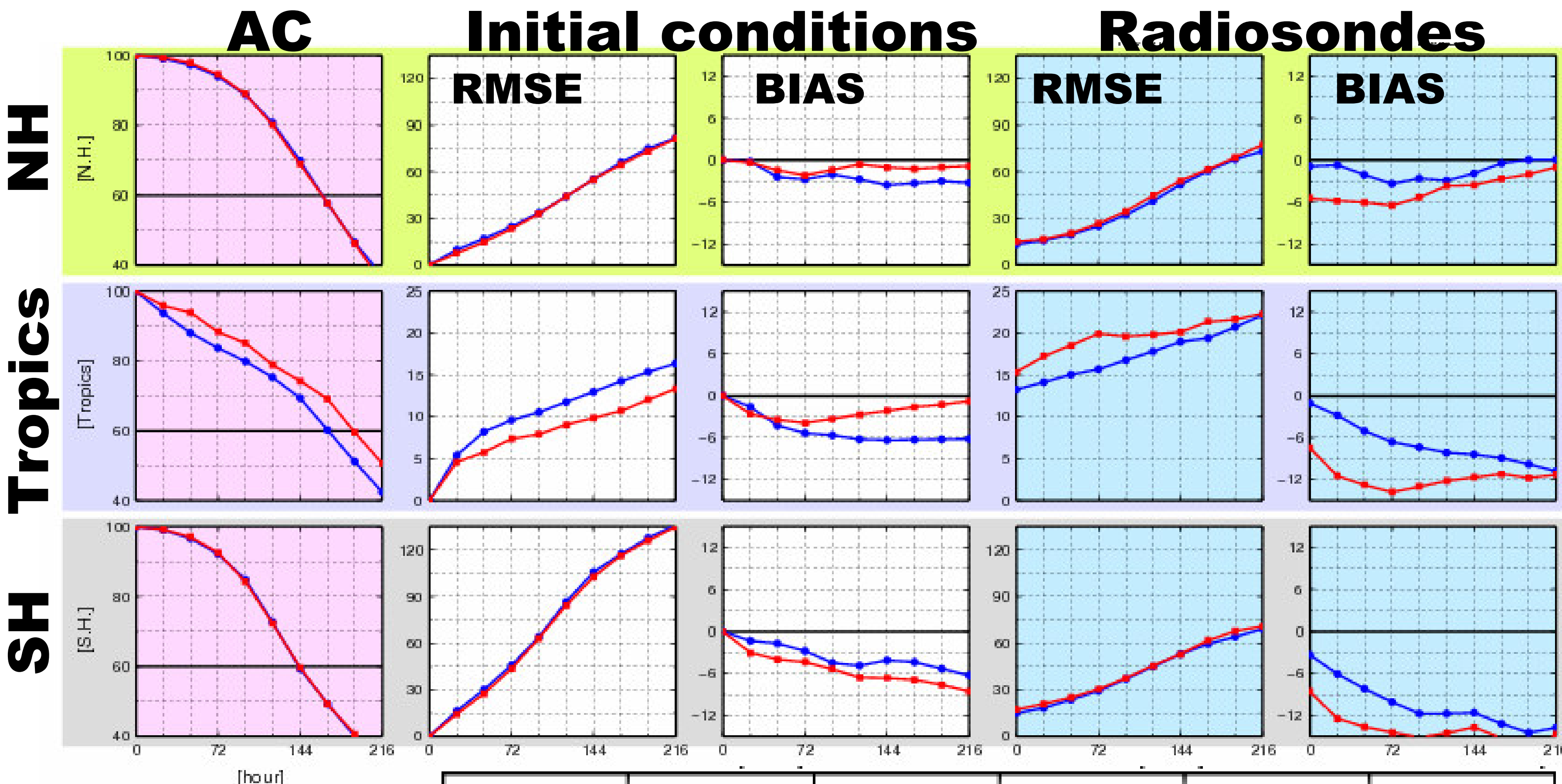
	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	–6.19	–4.36	–5.71	3.66	1.32
N. Hem.	–4.18	1.12	0.91	3.98	0.57
Tropics	6.86	3.39	3.09	14.07	10.21
S. Hem.	–7.60	–8.91	–7.91	–0.08	–1.62

Some bugs fixed in surface emissivity calculation

	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	–5.21	–2.33	–4.21	3.94	1.73
N. Hem.	–3.89	2.06	1.32	4.30	1.30
Tropics	7.05	6.49	7.44	13.58	9.57
S. Hem.	–6.35	–6.47	–6.20	0.39	–1.14

Period: August 2004

Comparison with 3D-Var



Red: LETKF
Blue: 3D-Var

Z500

	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	3.62	3.07	4.67	9.55	7.60
N. Hem.	0.32	4.32	5.28	7.93	5.09
Tropics	15.23	15.15	22.76	21.33	18.00
S. Hem.	3.84	0.23	4.00	6.08	5.41

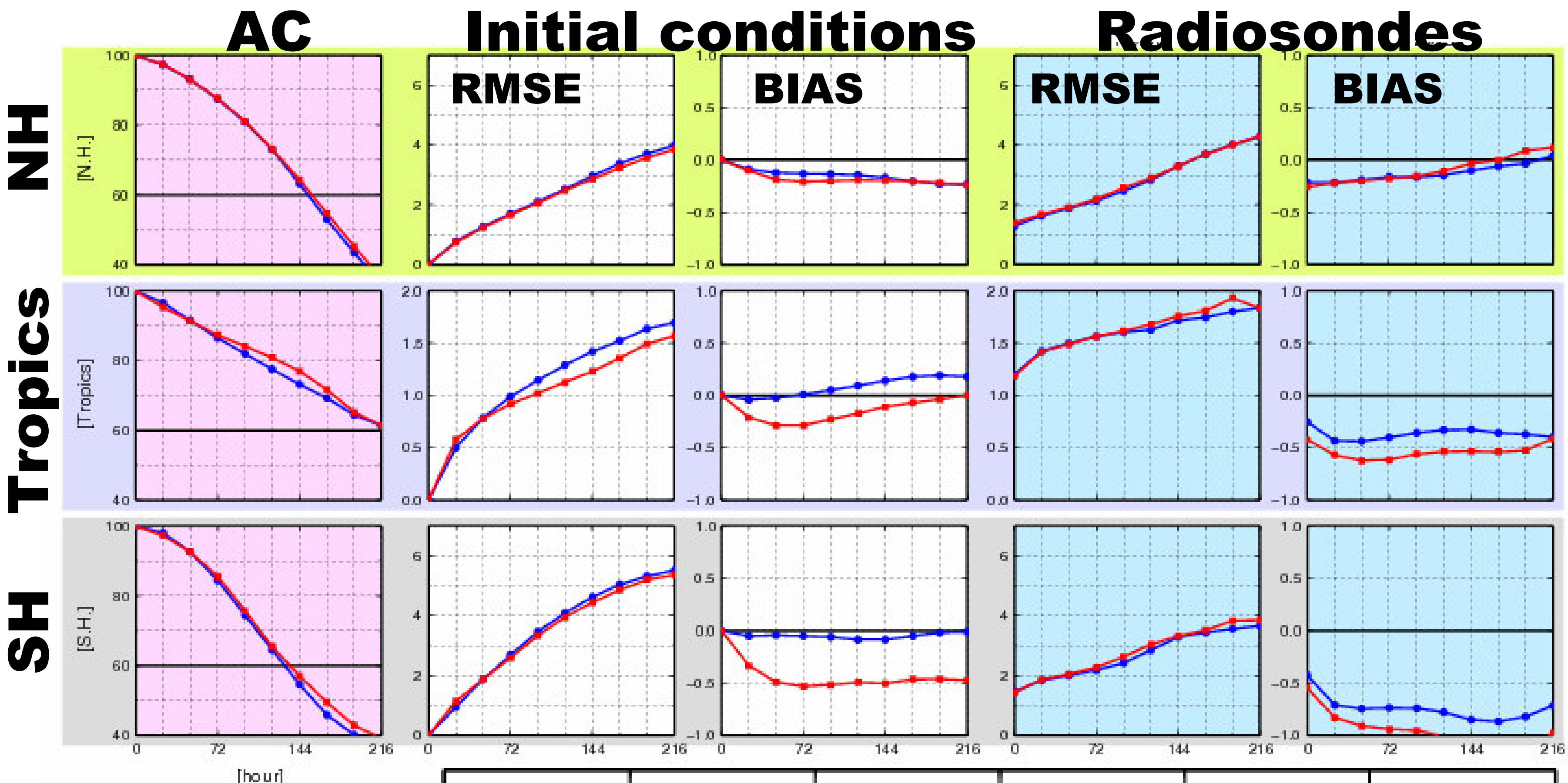
Period: August 2004

Reason for bias drifts in AMSU-A 4ch

FACTS:
✓ 4D-Var uses RTTOV-7
✓ LETKF uses RTTOV-8
✓ AMSU-A ch.4 is sensitive to surface emissivity and lower tropospheric temperature
✓ A known bug in the surface emissivity model "FASTEM-2" in RTTOV-7, where the surface emissivity is spuriously overestimated

- 4D-Var VarBC corrects the "spurious" bias caused by the bug
- Therefore, observed radiances (bias corrected) are too large for LETKF
- Thus, the lower troposphere is heated by assimilating the too large radiance observations, which explains the cold forecast bias relative to analysis (because analysis is too warm)
- The adaptive BC within LETKF corrects the wrong bias

Experiments without satellite radiances



Red: LETKF
Blue: 4D-Var

T850

9–days	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	2.68	1.69	2.10	6.60	4.30
N. Hem.	–0.72	2.98	3.09	5.88	3.22
Tropics	8.45	6.19	1.13	14.43	11.60
S. Hem.	3.19	0.52	2.10	4.66	2.34

Computational time

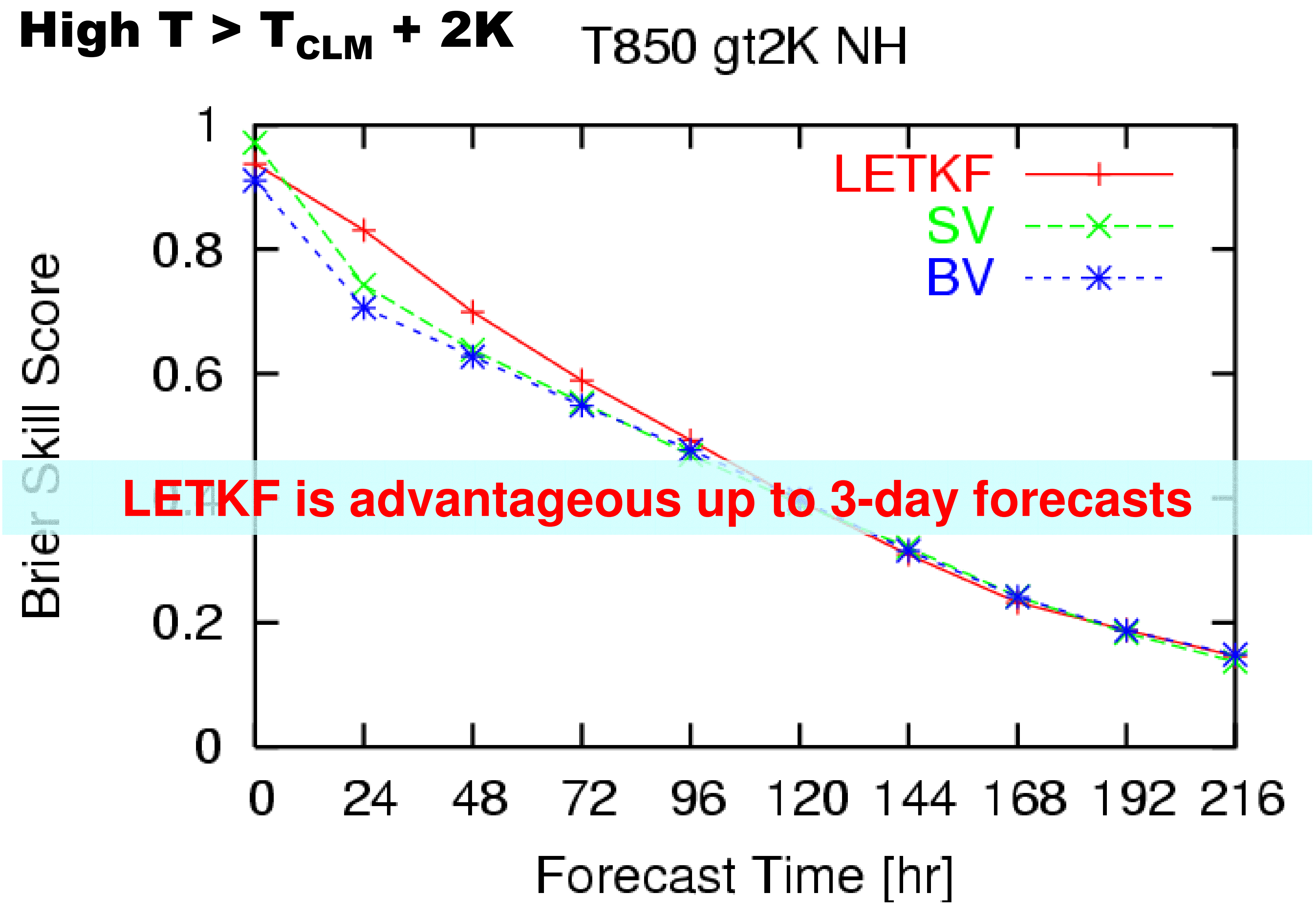
LETKF	4D-Var
11 min x 60 nodes 5 min for LETKF 6 min for 9-hr ensemble forecasts	17 min x 60 nodes
TL319/L60/M50	Inner: T159/L60 Outer: TL959/L60

Estimated for a proposed next generation operational condition

6 min (measured) x 8 nodes for LETKF with TL159/L40/M50

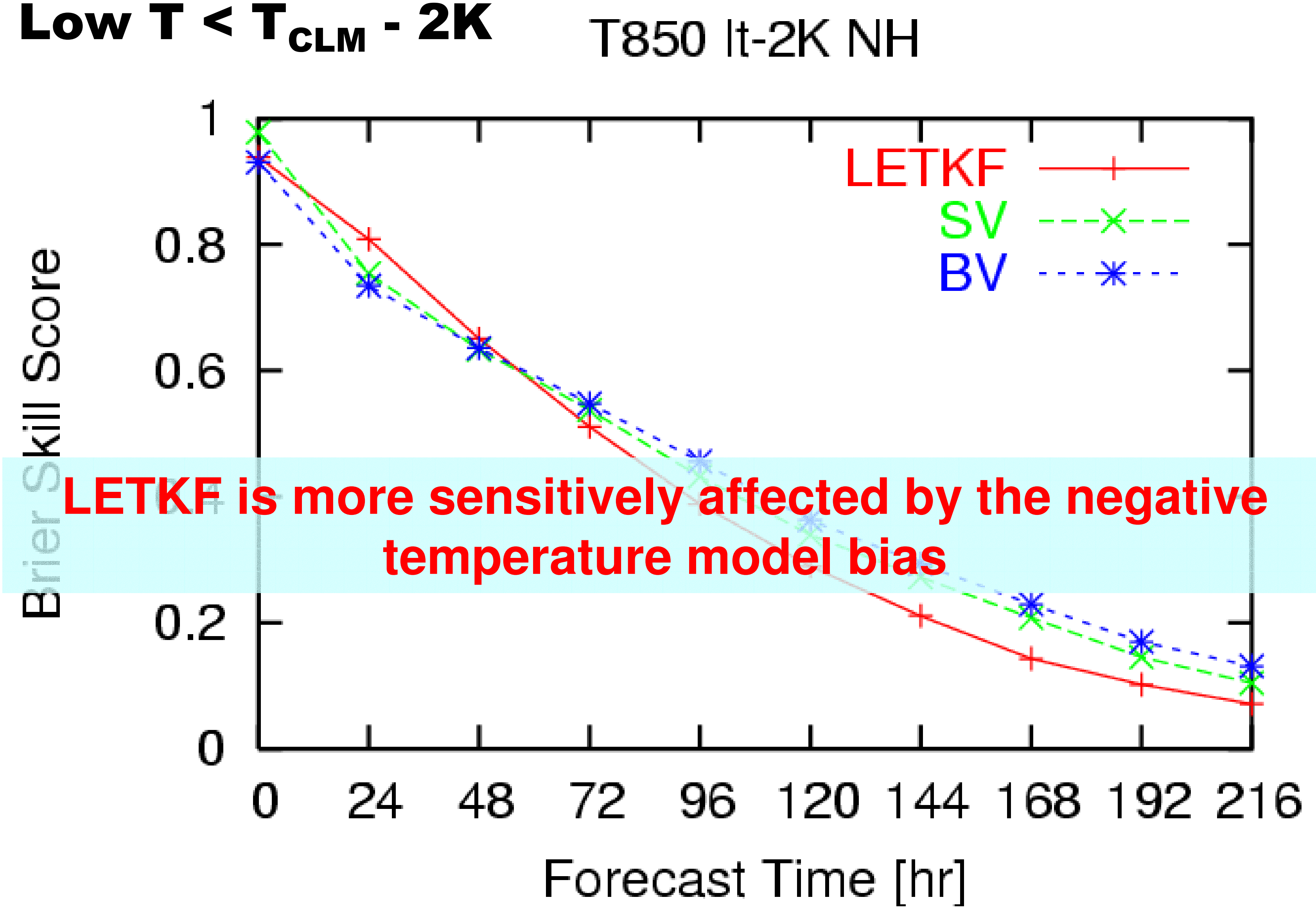
Computation of LETKF is reasonably fast,
good for the operational use.

Probabilistic forecast skills



Probabilistic forecast skills

Low T < T_{CLM} - 2K



Summary

- Comparison with 3/4D-Var:

NH	LETKF ~ or > 4D-Var >> 3D-Var
Tropics	LETKF >> 4D-Var >> 3D-Var
SH	4D-Var >> LETKF > 3D-Var

- LETKF is advantageous in Typhoon prediction
- Probabilistic forecast skill is generally improved
  - Still a bias problem exists
- Need improvements in the treatment of satellite radiances

Future plan, ongoing development

- Better use of satellite radiances
  - QC system with RTTOV-8
    - Collaborating with satellite scientists

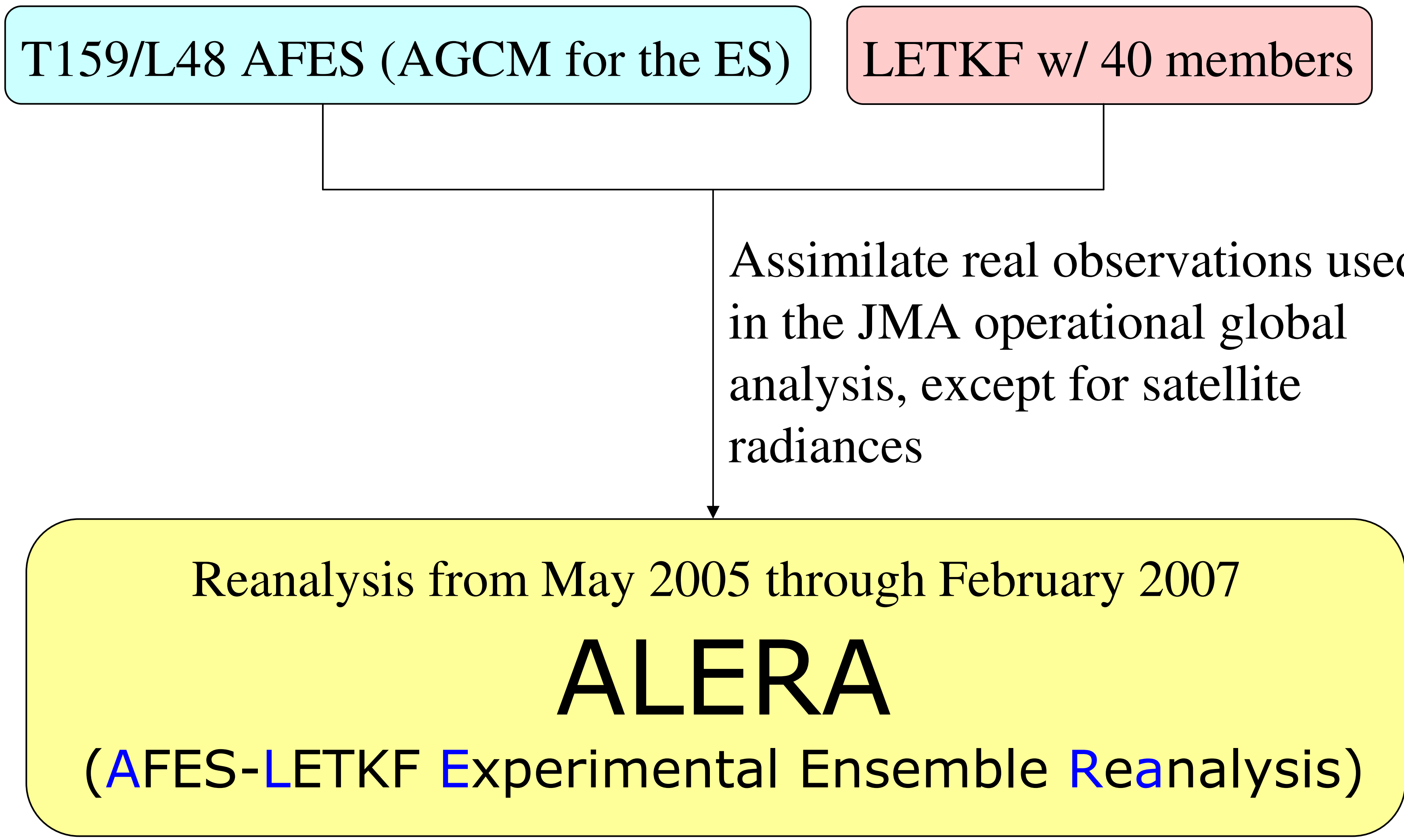
- Adapting to a higher resolution
  - Next-generation global model with the reduced Gaussian grid at a TL319/L60 resolution
  - kd-tree search algorithm to select local obs (geographical range search) will be implemented

cf. Eric Kostelich already applied kd-tree at UMD

Make comparisons with the next-generation operational system

Researches with the Earth Simulator

Ensemble Reanalysis: ALERA



ALERA dataset

ALERA

(AFES-LETKF Experimental Ensemble Reanalysis)

data are now available online for free!!

http://www3.es.jamstec.go.jp/

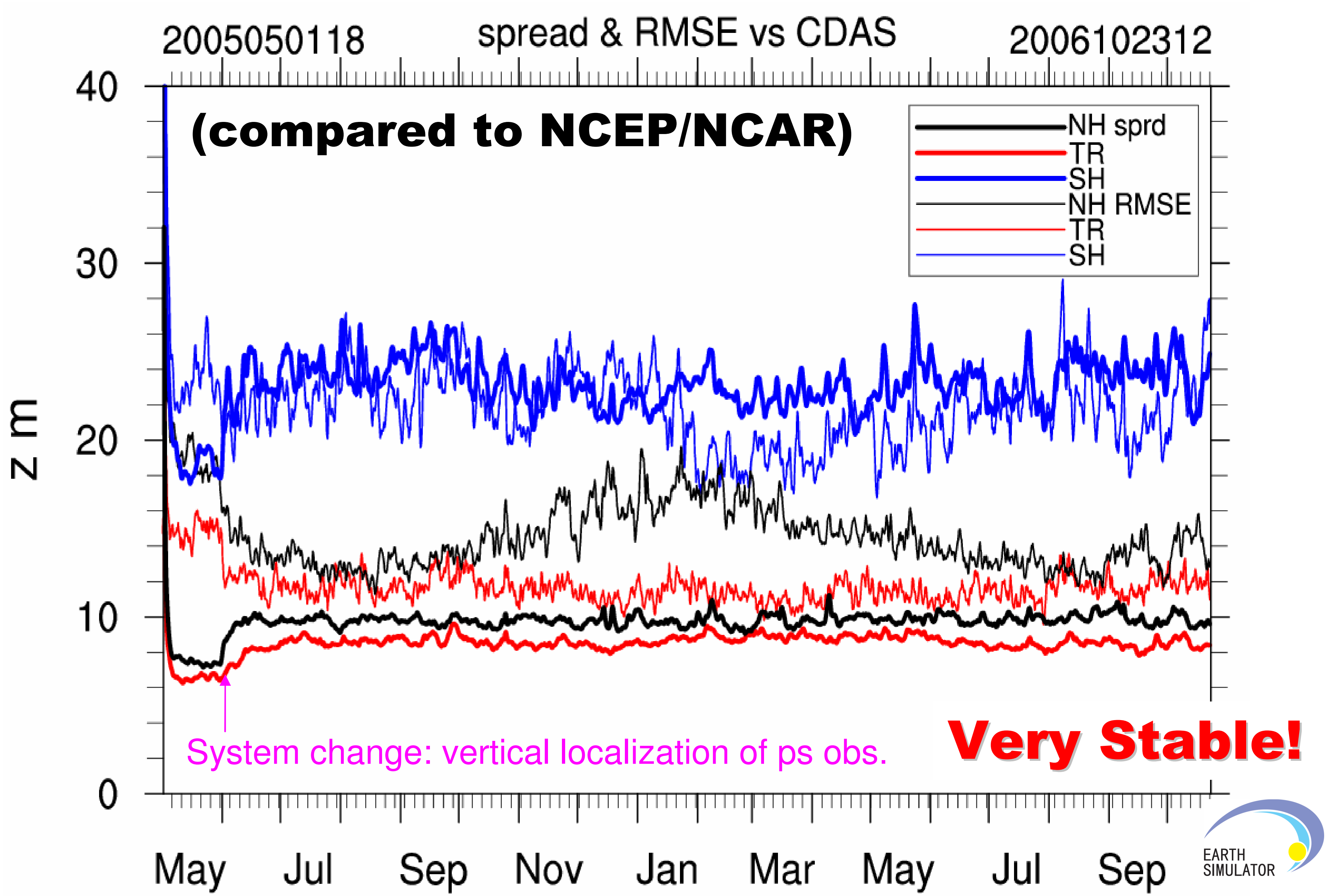
Contents

Ensemble reanalysis dataset for over 1.5 years since May 1, 2005

- 40 ensemble members
- ensemble mean
- ensemble spread

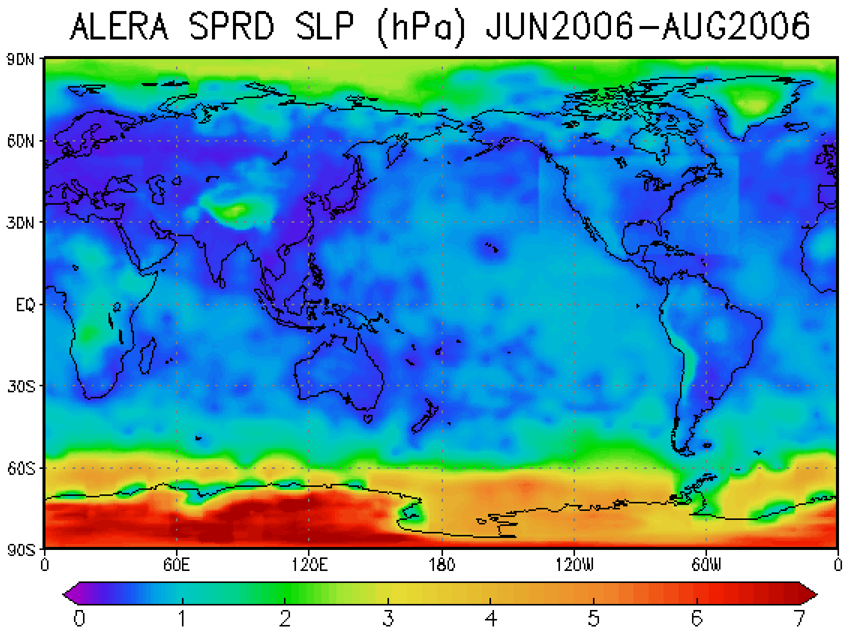
Available 'AS-IS' for free ONLY for research purposes
Any feedback is greatly appreciated.

Stable performance

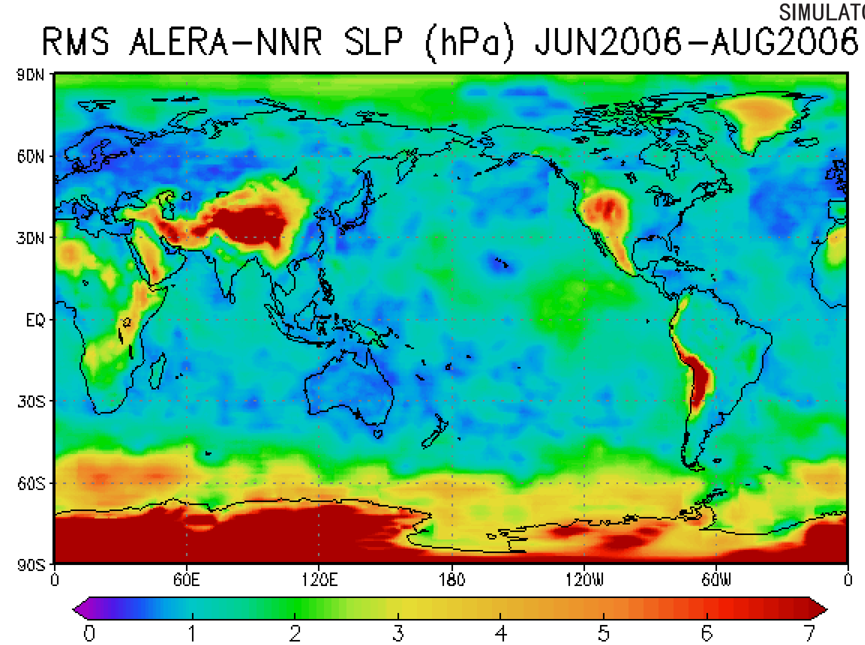


Ensemble spread and RMS diff

SPREAD



RMS Diff

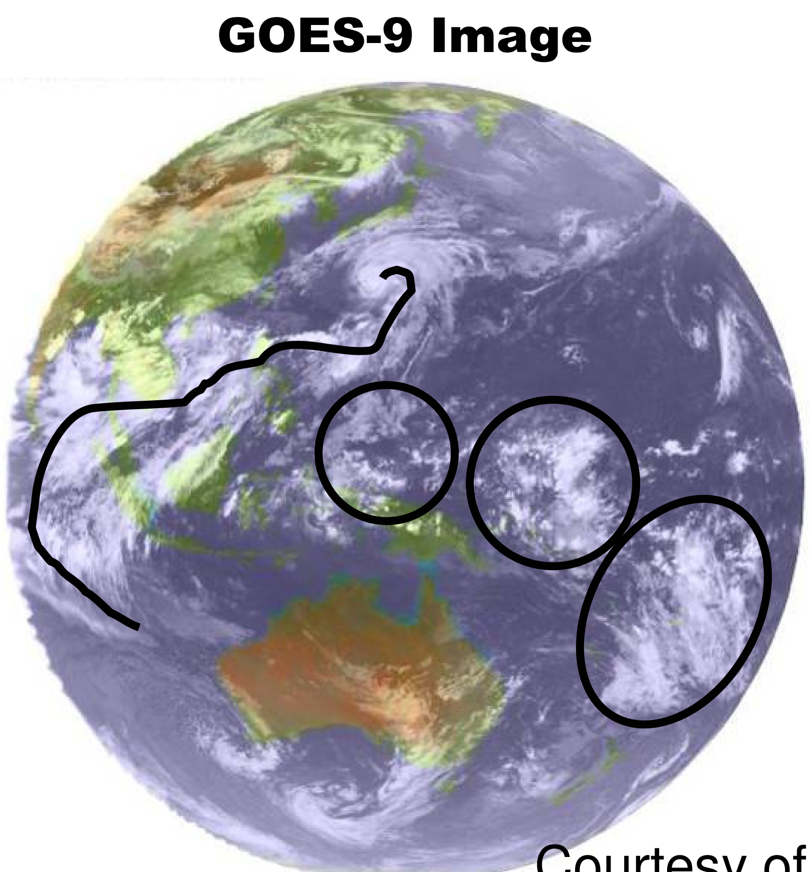


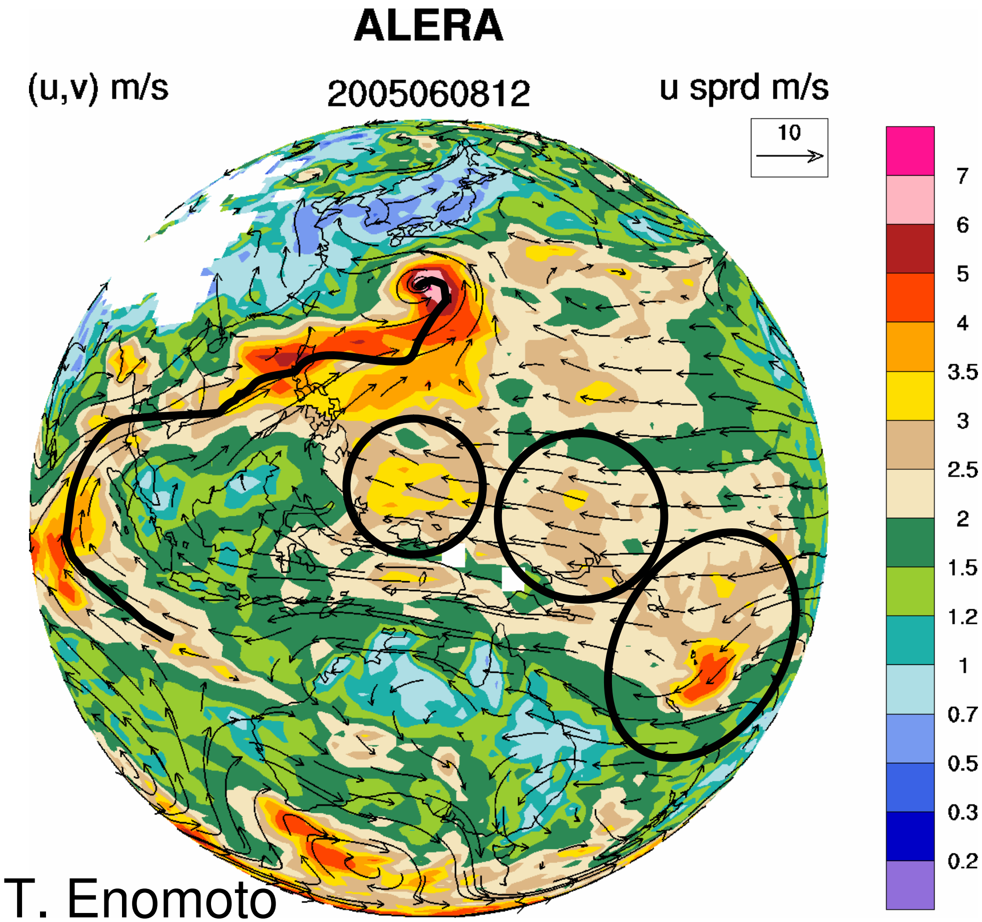
- Generally similar pattern
- Spread may need to be calibrated

  (Spread should be smaller since NNR contains errors)
- Underestimated spread by dense observations

Analyzing the analysis errors

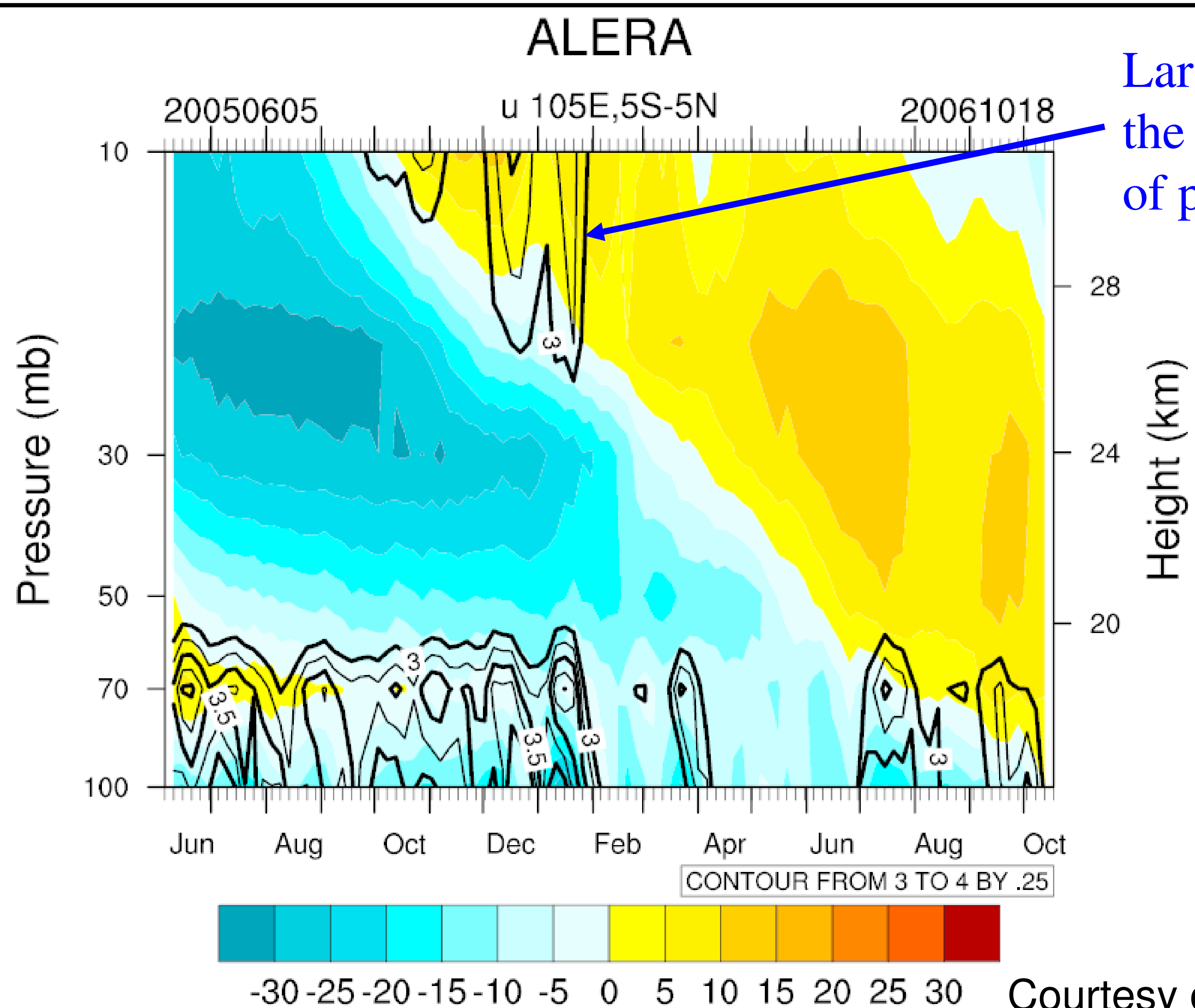
- EnKF provides not only analysis itself but also the analysis errors (or uncertainties of the analysis)
- What is the dynamical meaning of the analysis errors?





Courtesy of T. Enomoto

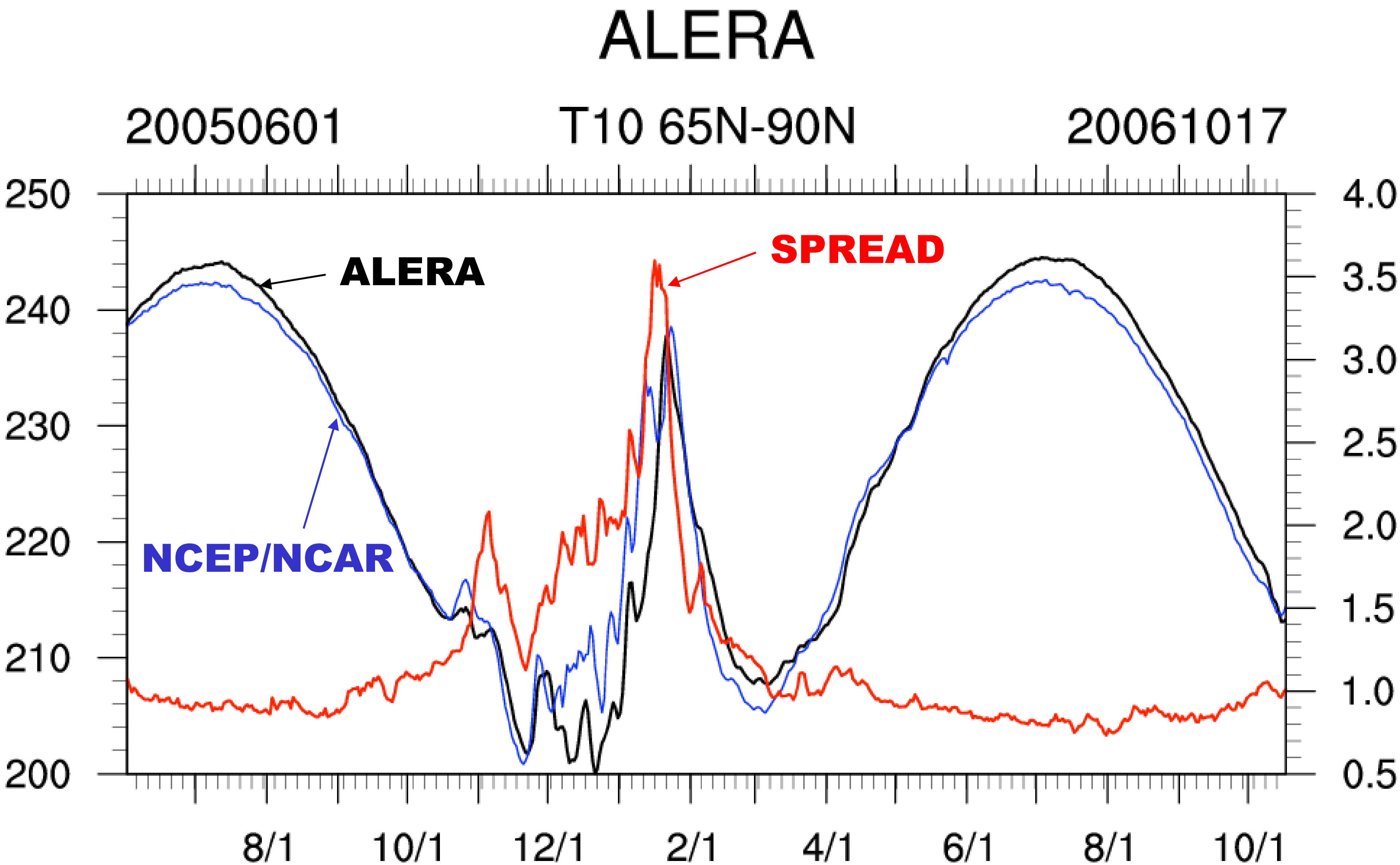
QBO and ensemble spread



Large spread at the initial stage of phase change

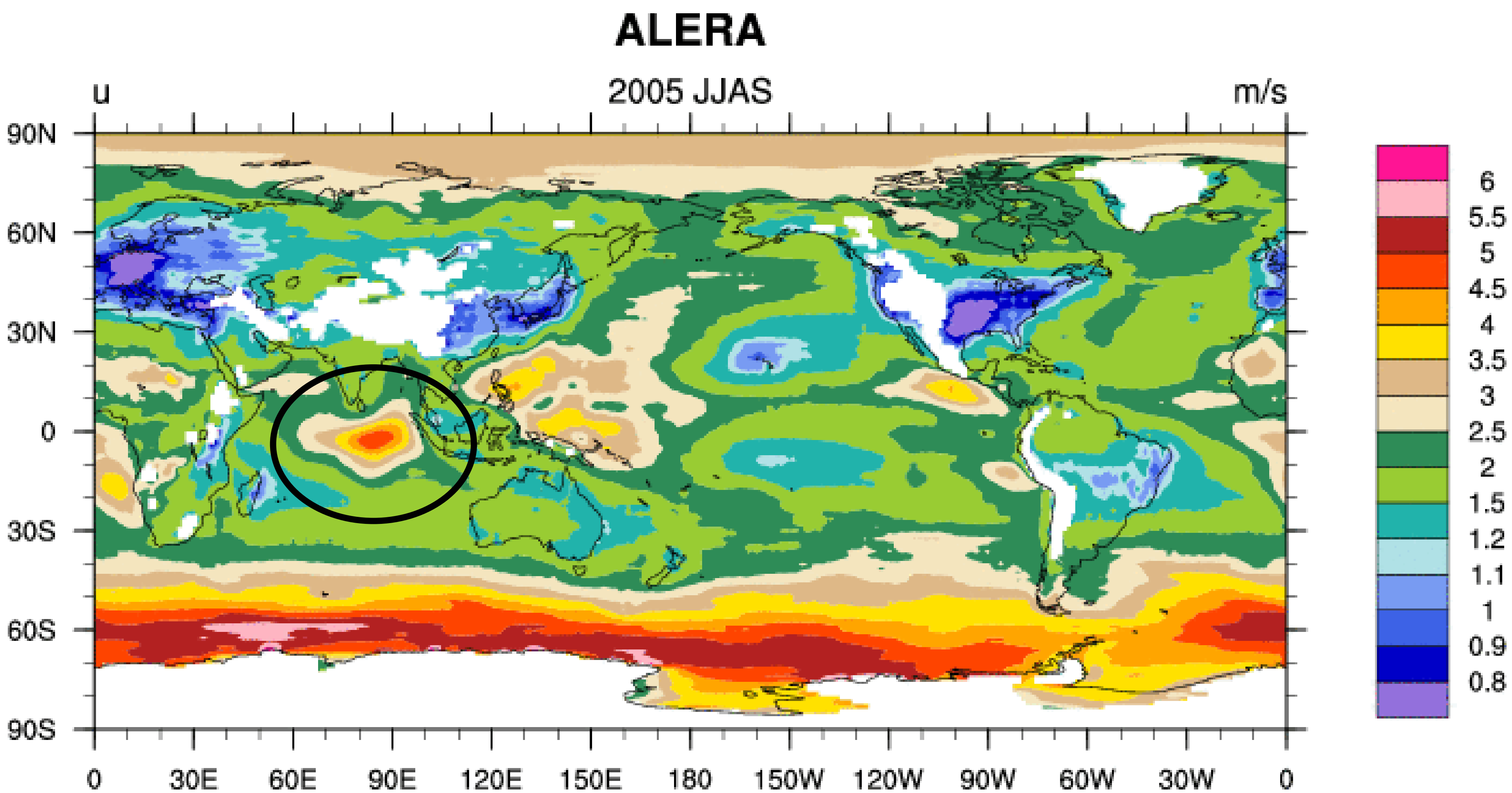
Courtesy of T. Enomoto

Stratospheric sudden warming



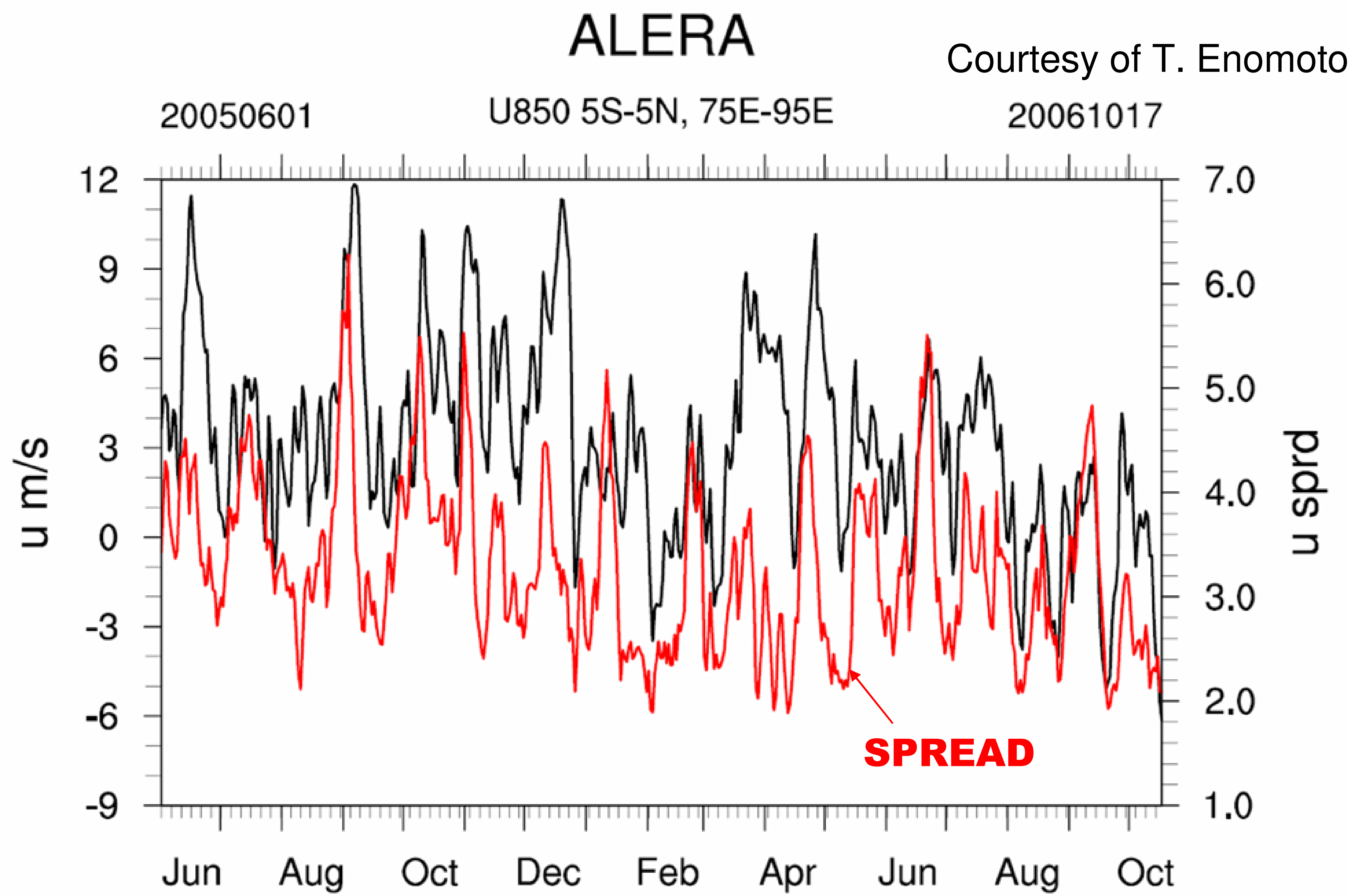
Courtesy of T. Enomoto

Large spread in tropical lower wind

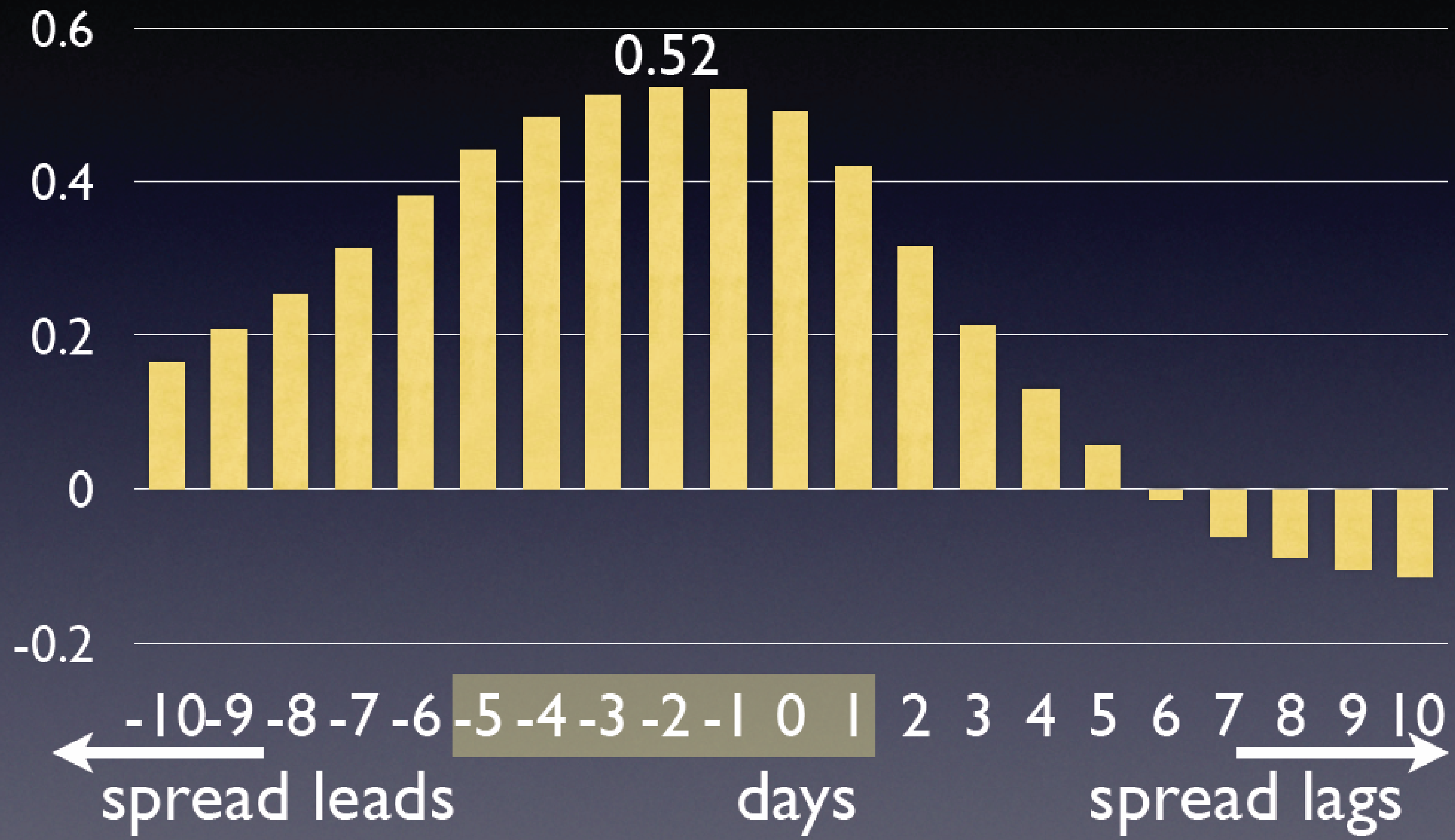


Courtesy of T. Enomoto

Tropical lower wind and the spread



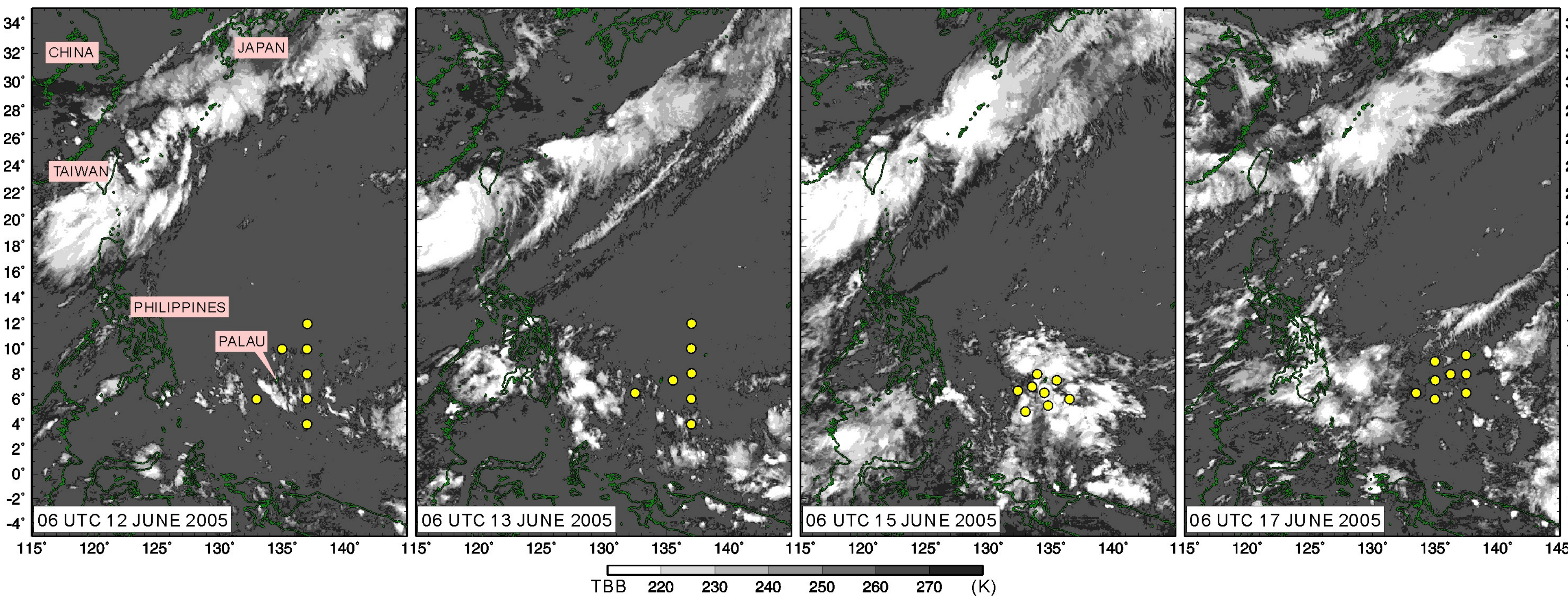
Lag correlation between mean and spread



Courtesy of T. Enomoto

Collaboration with observing scientists

There was an intensive observing project in the tropical western Pacific (a.k.a. PALAU-2005); the dropsonde obs during June 12-17, 2005 have been assimilated with the AFES-LETKF system.



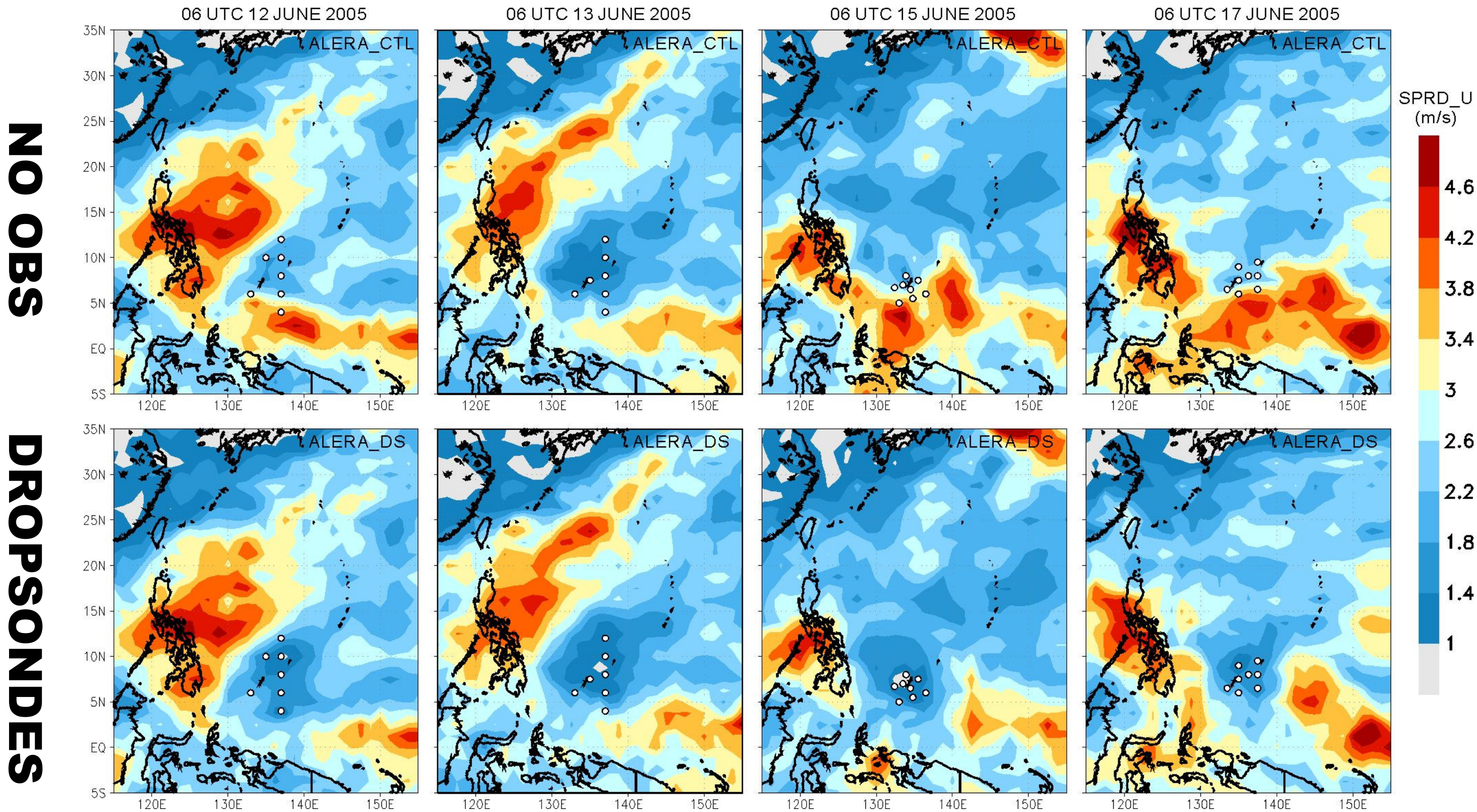
Satellite image and dropsonde locations

Moteki et al., 2007: *SOLA*

Collaboration with observing scientists

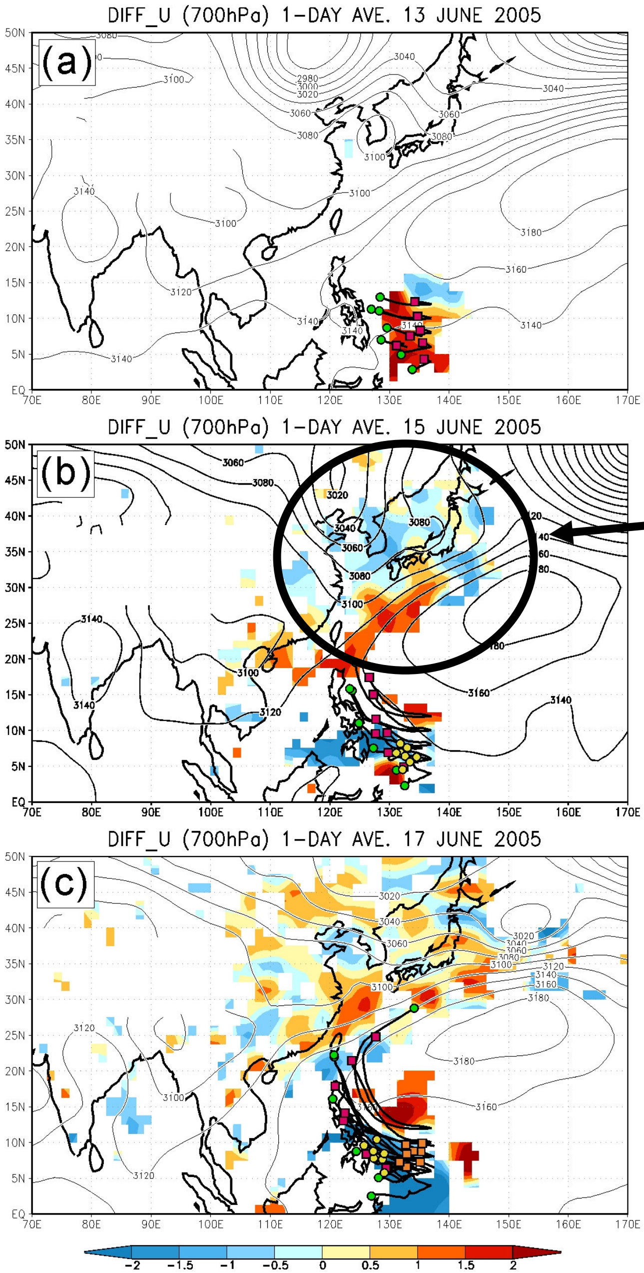
Impact by dropsonde observations

Analysis ensemble spreads by LETKF



Moteki et al., 2007: *SOLA*

Collaboration with observing scientists



Propagation of observing signals

cf. Szunyogh et al. (2000; 2002)
Hodyss and Majumdar (2007)

Faster propagation than the advection speed, the faster speed (~12 m/s) corresponds to Rossby wave propagation

Moteki et al., 2007: *SOLA*

Summary

- Ensemble spread represents errors well.
- There seems to be dynamical meanings of analysis ensemble spread, which could be investigated in various scales.
- Collaborative study with observing scientists has just begun.

Research plans with the ES

- ALERA-2
  - 5-yr reanalysis (21st century reanalysis)
  - AFES-MATSIRO (SiB) (atmos-land coupled)
  - More diagnostics (OLR, Precipitation, land variables, etc.; any requests?)

- CFES-LETKF
  - LETKF with coupled atmos-land-ocean model

- More observing projects (e.g., PALAU-2008)

Other ongoing activities with LETKF

- Collaboration with chemical/aerosol-transport modeling scientists
  - Dr. Thomas Sekiyama at the Met Research Institute, JMA
  - Dr. Nich Schutgens at the Center for Climate System Research, University of Tokyo

- Collaboration with university students and scientists to assimilate atmospheric lidar wind data into a nonhydrostatic fine-mesh model
  - Prof. Toshiki Iwasaki at the Tohoku University

Collaborators

AFES-LETKF

- Prof. Shozo Yamane (Chiba Institute of Science and FRCGC/JAMSTEC, AFES)
- Dr. Takeshi Enomoto (ESC/JAMSTEC, AFES)
- Dr. Qoosaku Moteki (IORGC/JAMSTEC, Observing scientist)

NHM-LETKF

- Kohei Aranami (NPD/JMA, NHM)
- Dr. Hiromu Seko (MRI/JMA, DA with NHM)

GSM-LETKF

- Yoshiaki Sato (NPD/JMA, staying at NCEP)
- Takashi Kadowaki (NPD/JMA, 4D-Var)
- Ryota Sakai (NPD/JMA, EPS)
- Munehiko Yamaguchi (NPD/JMA, Typhoon EPS)

Chemical Transport

- Dr. Thomas Sekiyama (MRI/JMA, Chemical model)

Acknowledgments

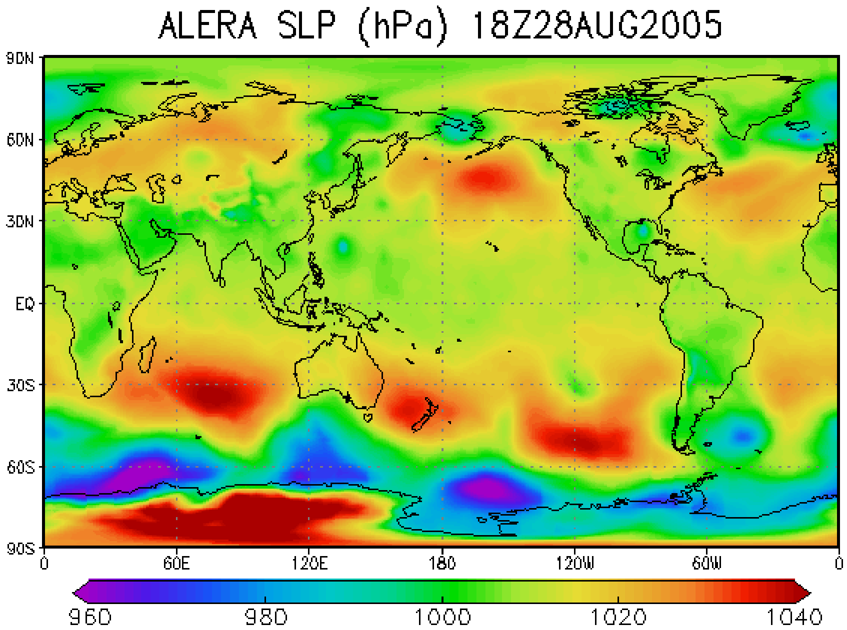
- The idea of LETKF has been developed at the University of Maryland.
- Radiative transfer model RTTOV-8 is provided by EUMETSAT.
- ALERA has been produced using the Earth Simulator.


Thank you
for your attention!

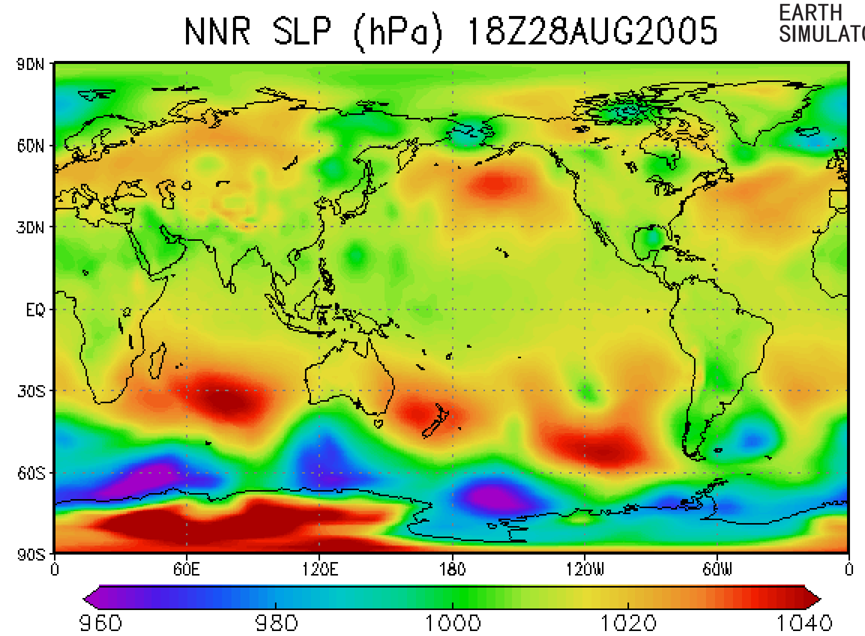
Supplement: ALERA accuracy

Snapshot (SLP)

ALERA



NNR

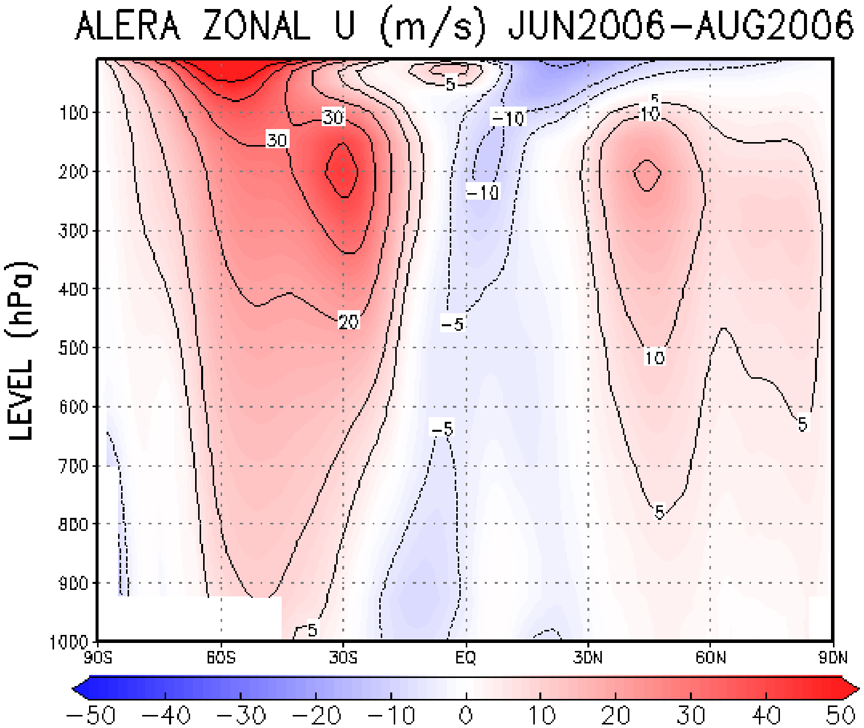


EARTH SIMULATOR

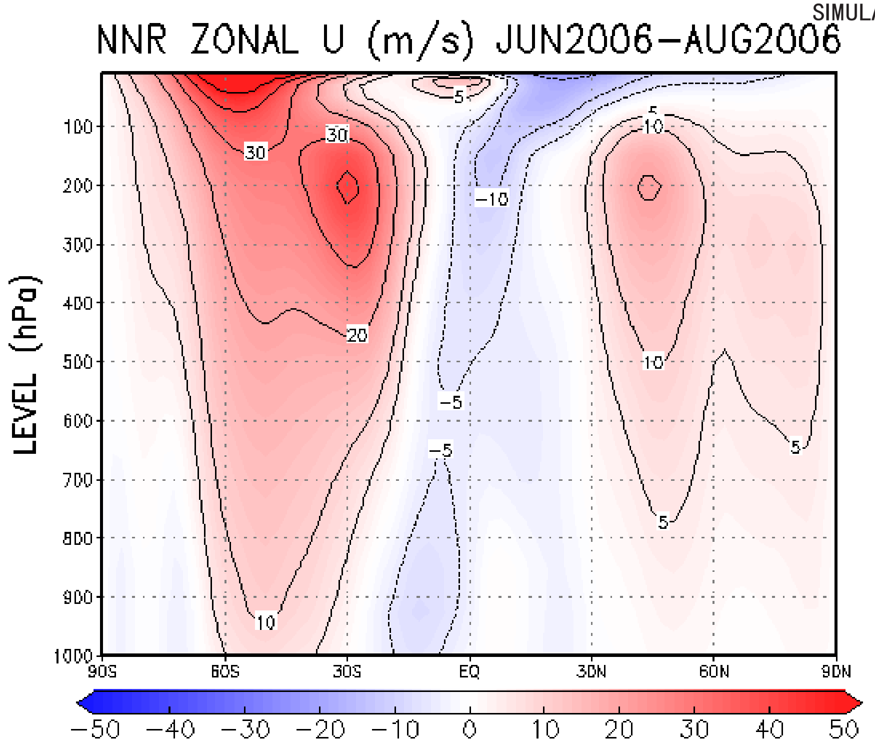
ALERA analysis is almost identical to NNR.

Zonal mean winds in JJA

ALERA

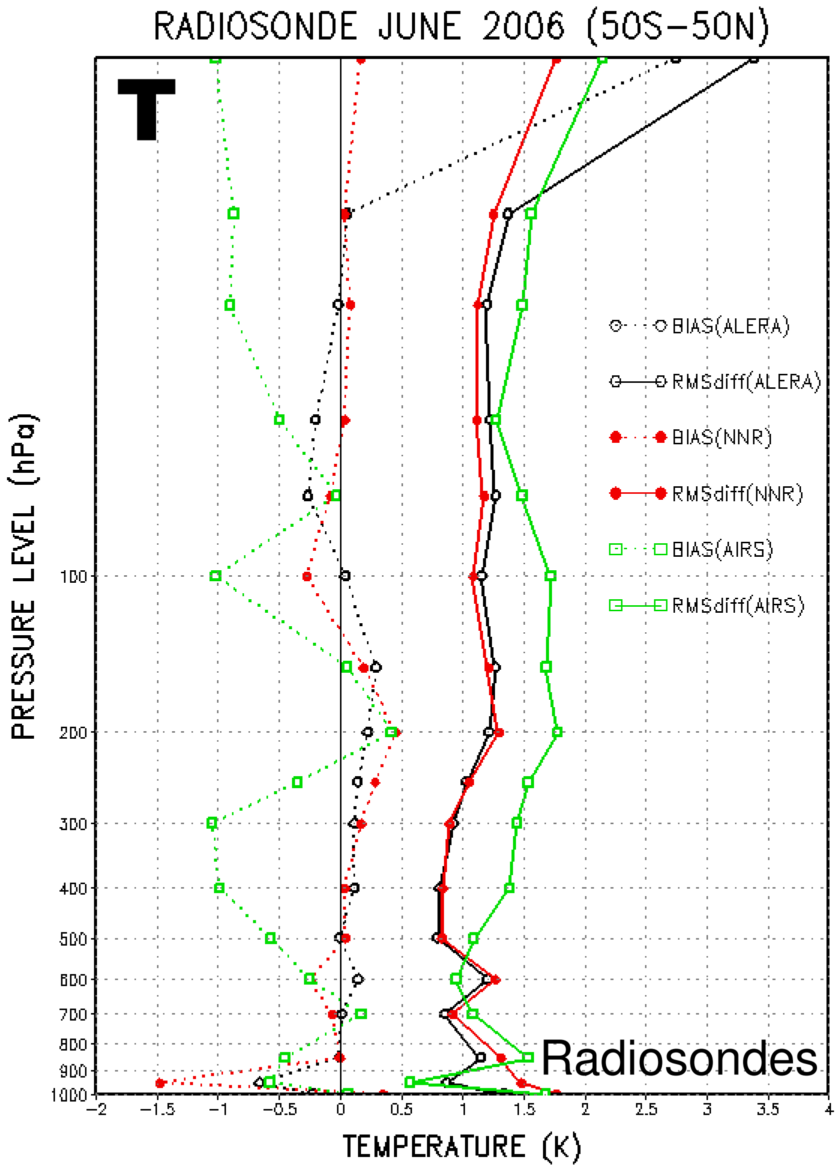


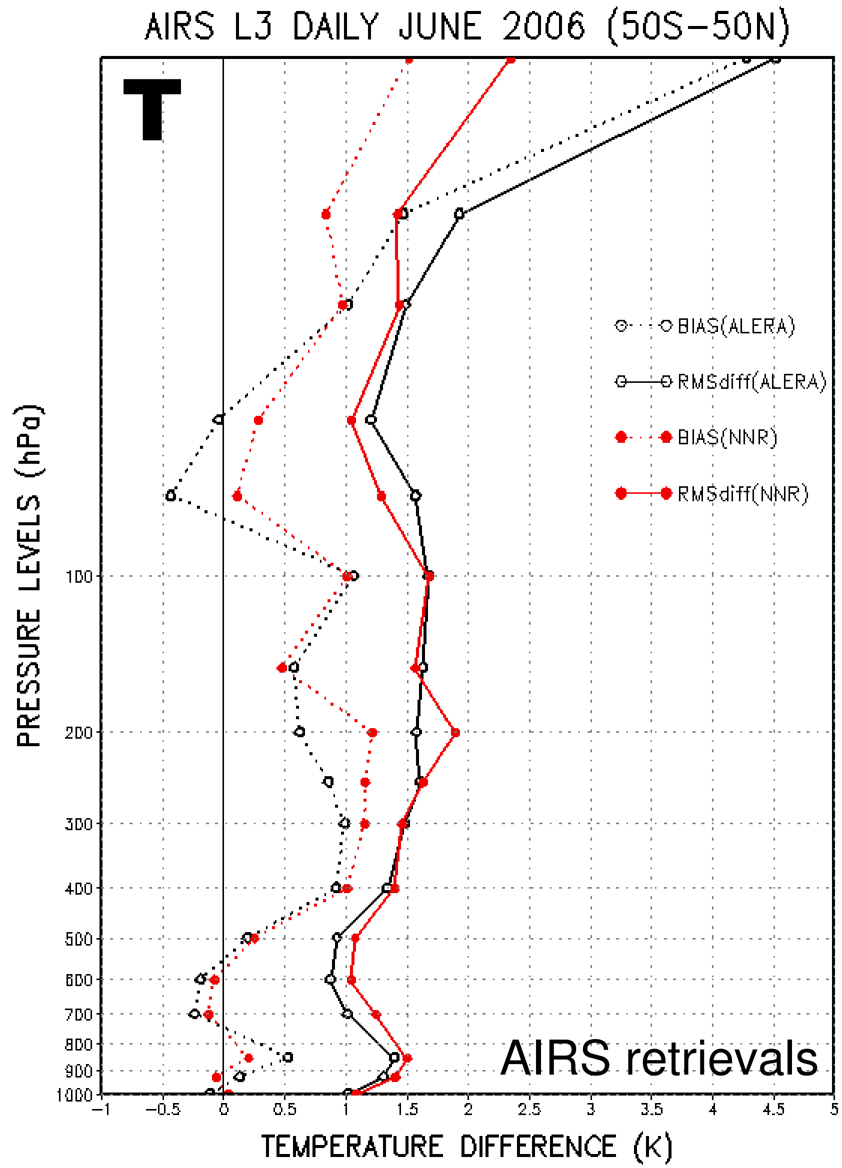
NNR



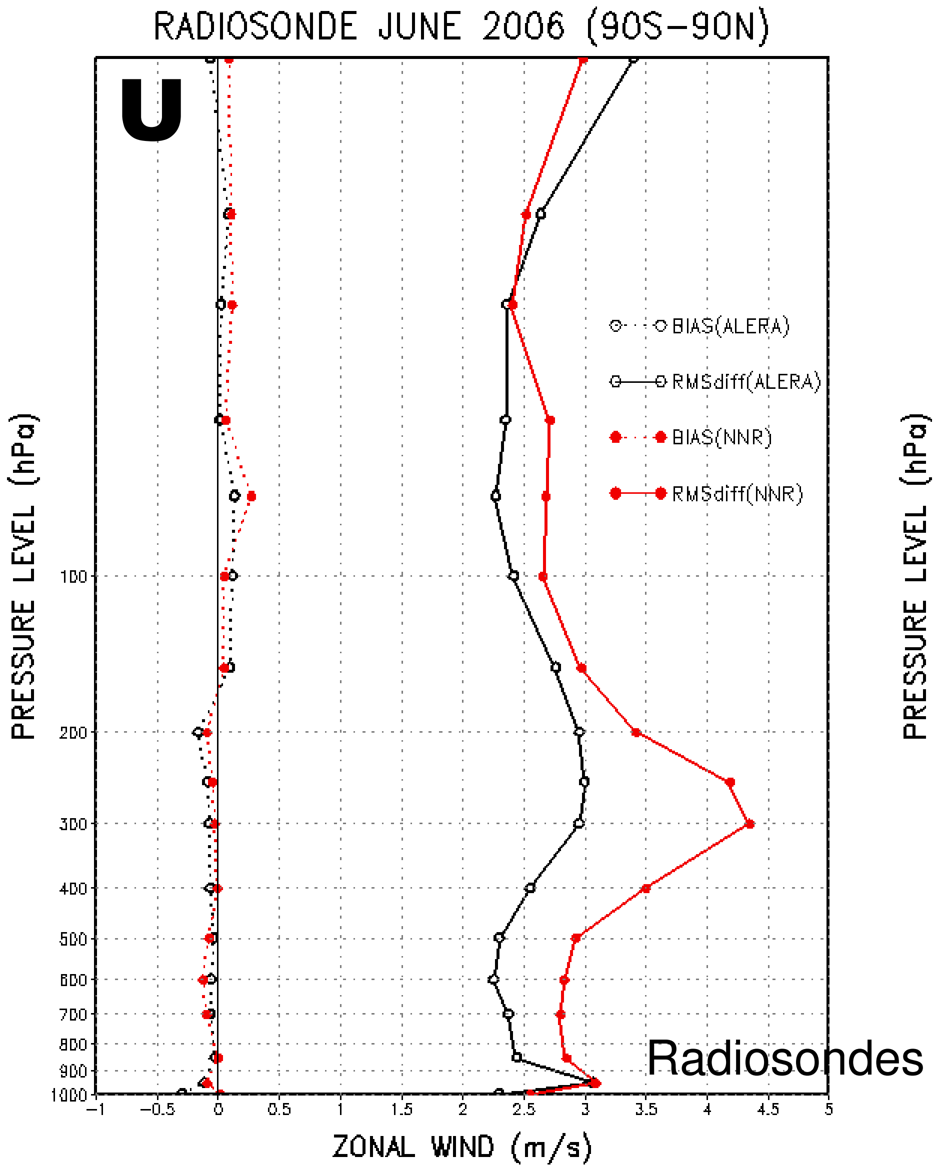
EARTH SIMULATOR

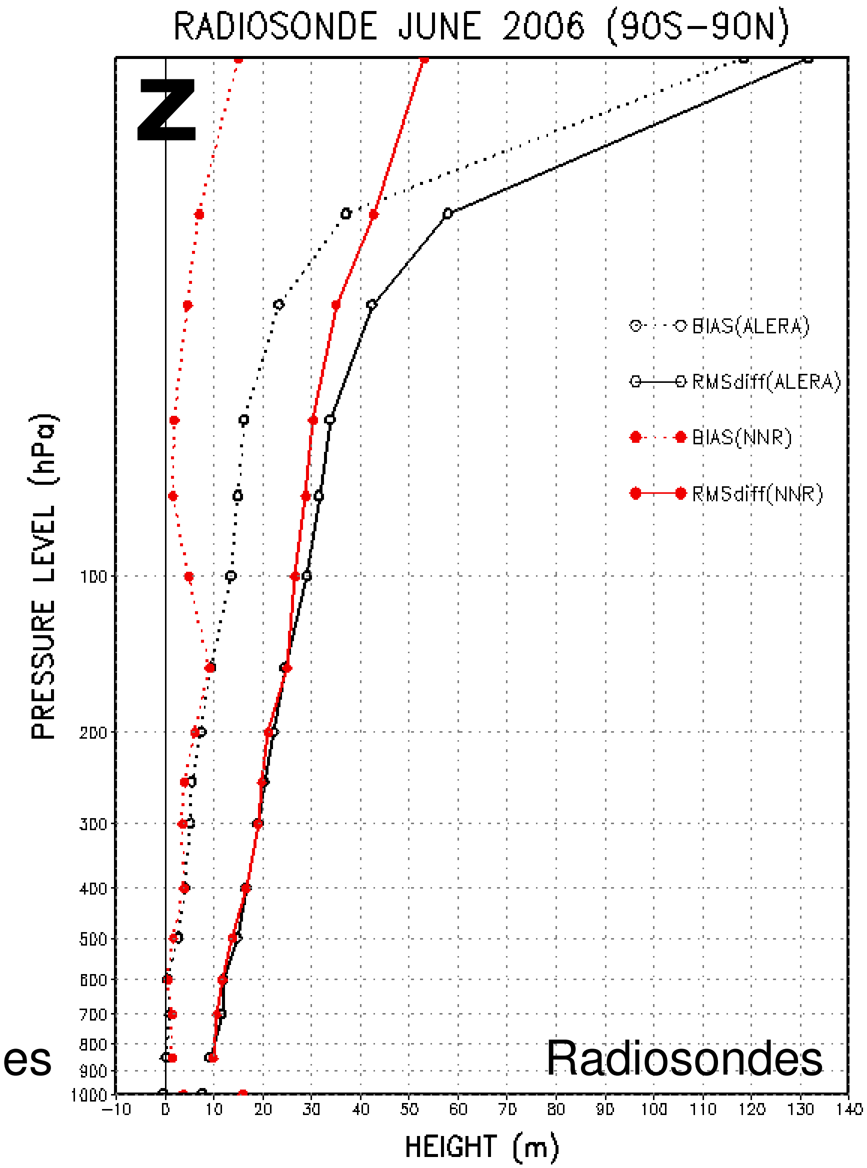
Comparison with observations



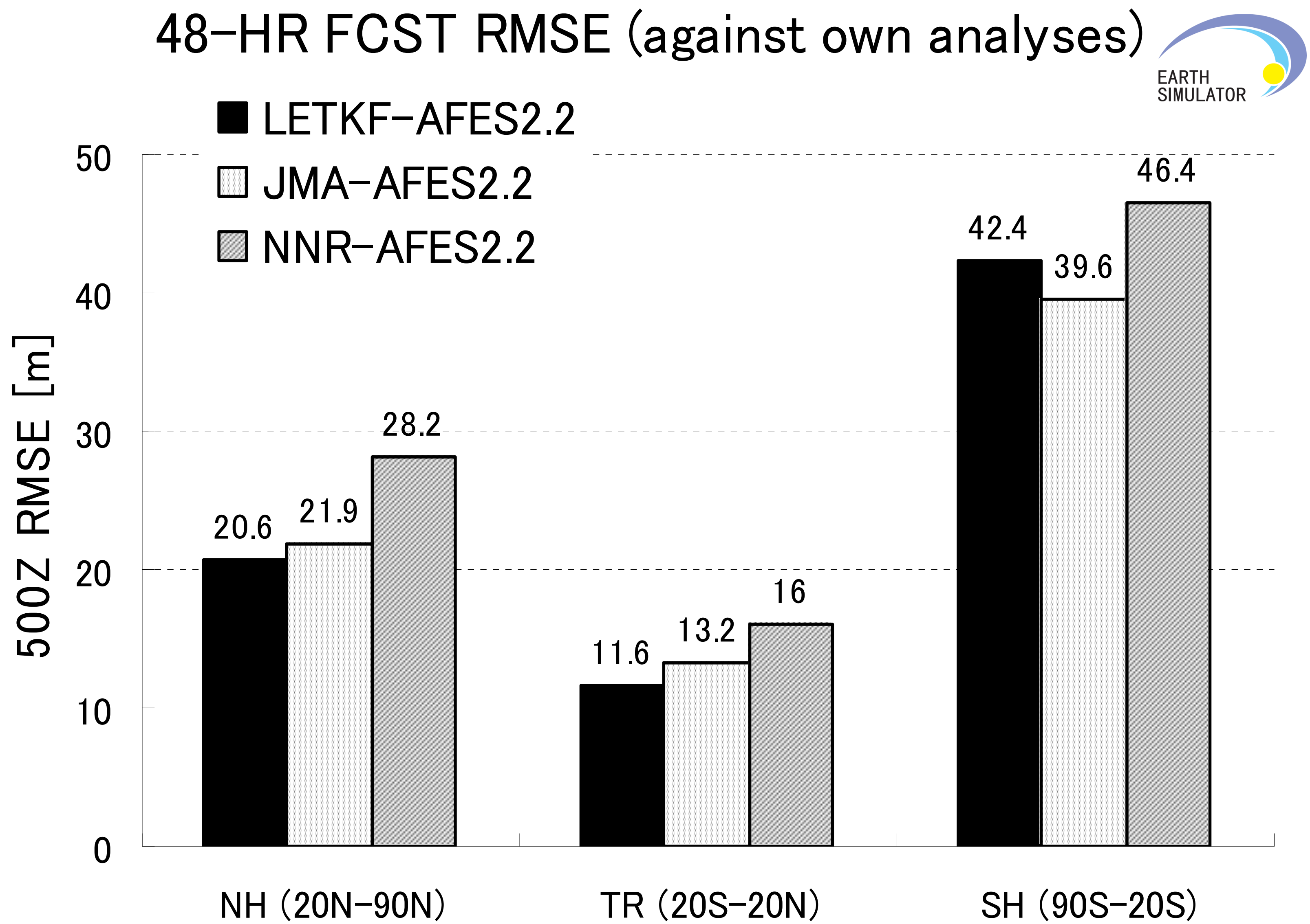


Compared with radiosondes





Forecast RMS errors



Adapted from Miyoshi and Yamane (2007)

Supplement: GSM-LETKF

Improvement (%) relative to 4D-Var

August 2004

	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	–2.53	–2.04	–1.57	5.51	3.56
N. Hem.	–2.77	2.16	2.09	4.91	1.63
Tropics	7.05	4.90	4.95	14.92	11.81
S. Hem.	–3.06	–5.49	–2.68	2.62	1.91

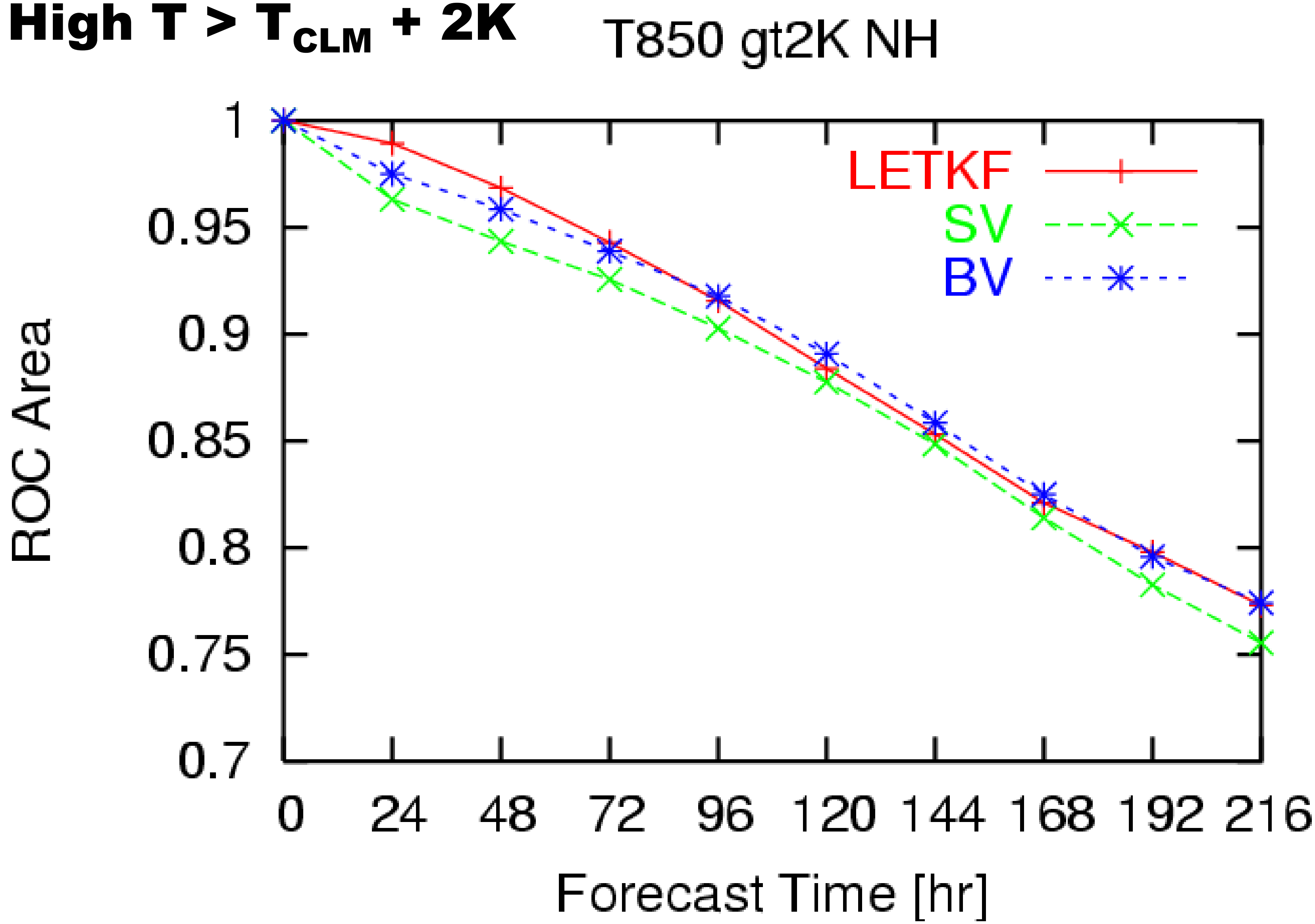
December 2005

	PseaSurf	T850	Z500	Wspd850	Wspd250
Global	–3.39	–1.27	–2.60	4.91	2.69
N. Hem.	–5.60	–0.29	–4.46	1.16	0.83
Tropics	8.74	4.94	13.43	16.52	12.10
S. Hem.	–1.06	–4.98	–0.82	3.79	0.15

LETKF is advantageous in the summer hemisphere

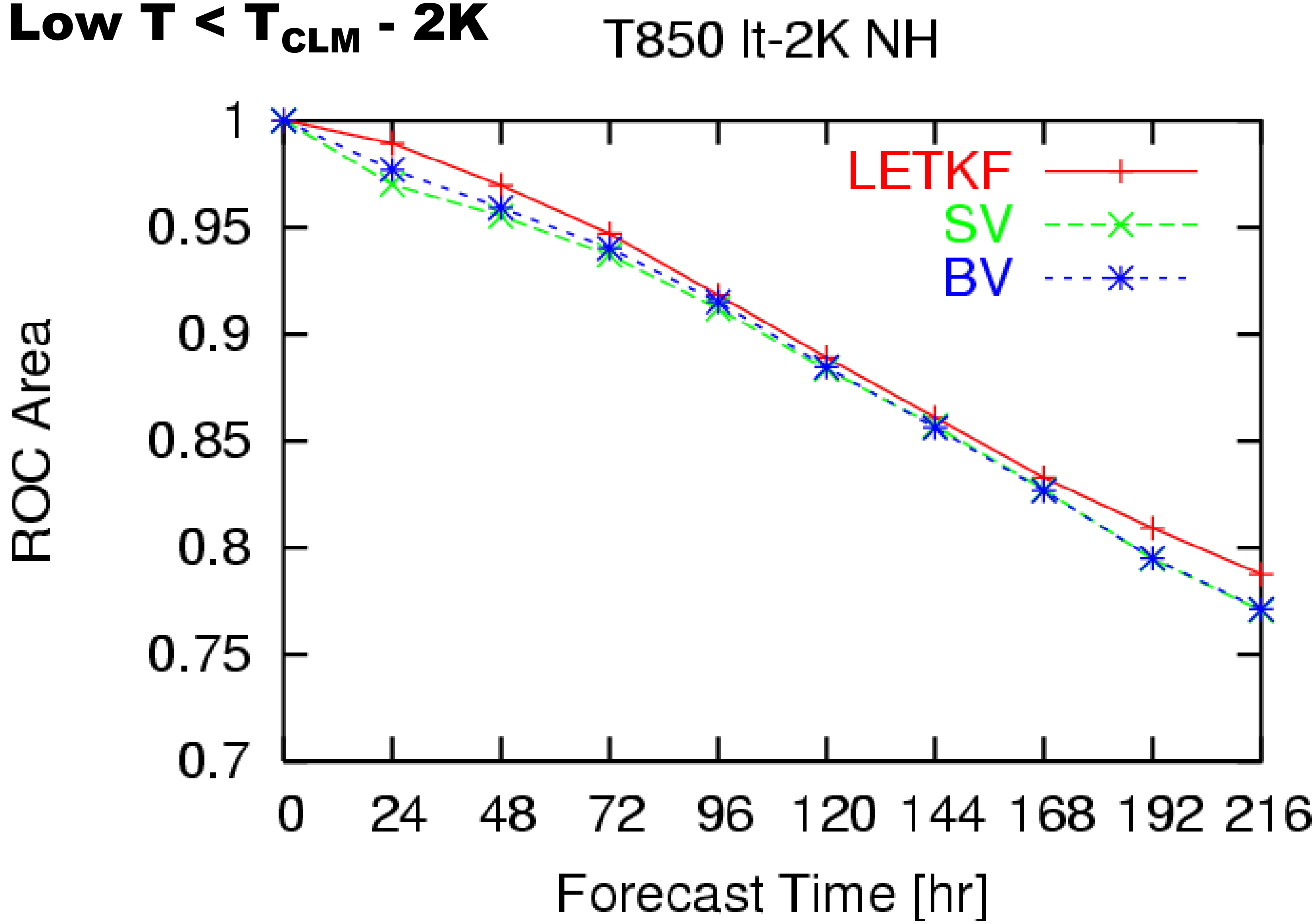
Economic value

High T > T_{CLM} + 2K

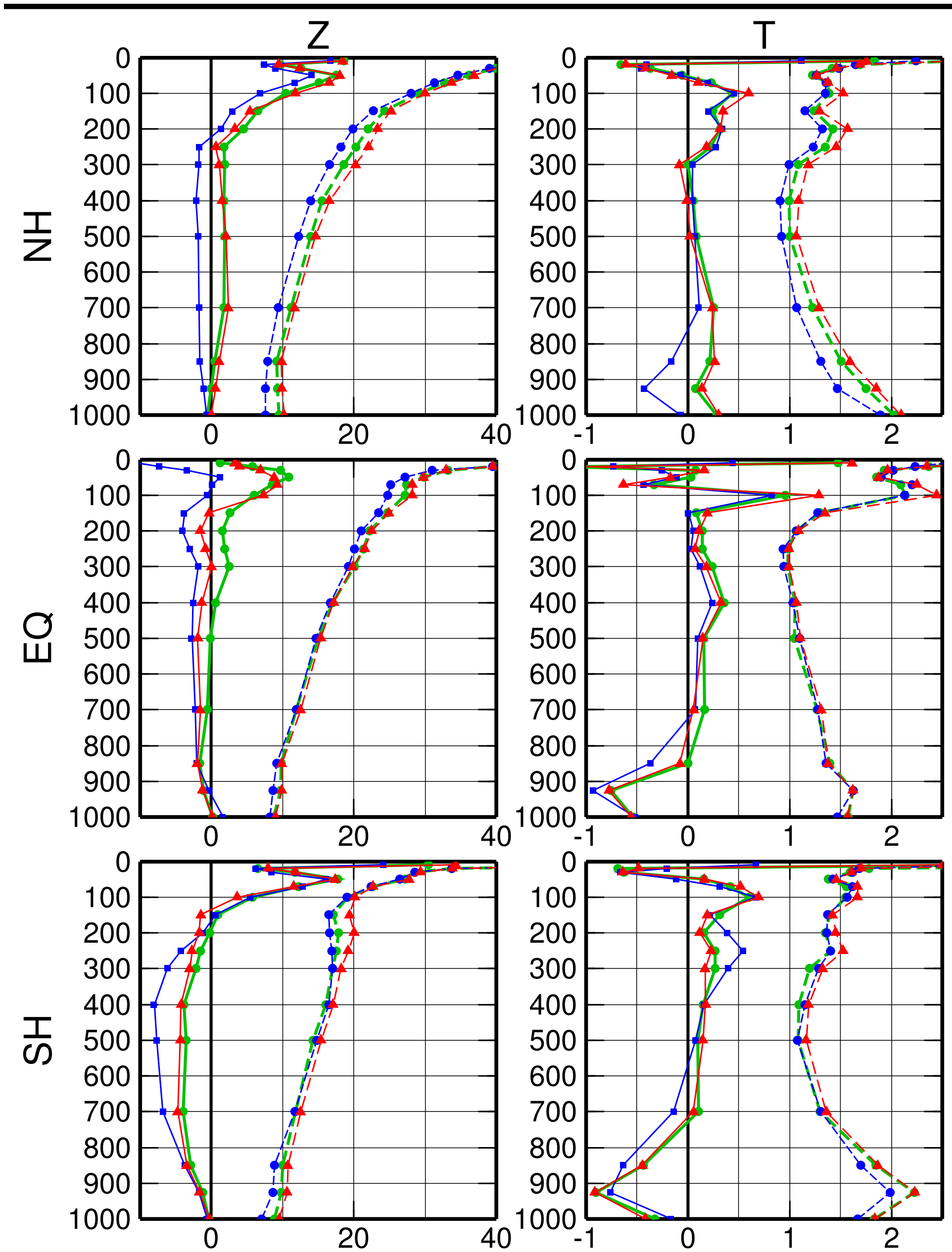


Economic value

Low T < T_{CLM} - 2K



Ensemble size 20 → 50



RMSE and bias against radiosondes

Blue: Operational 4D-Var

Red: 20-member LETKF

Green: 50-member LETKF

50 members > 20 members

Generally 4D-Var > LETKF

Exception: mid-upper tropospheric temperature in the SH

w/ satellite radiances