

Data Assimilation toward Big Data & Post-peta-scale Supercomputing: A Personal Perspective



Takemasa Miyoshi

RIKEN Advanced Institute for Computational Science

Takemasa.Miyoshi@riken.jp



With many thanks to

JMA

UMD Weather-Chaos group

JST CREST “Big Data Assimilation” project

JAXA PMM “Ensemble Data Assimilation” project

Data Assimilation Research Team

Data Assimilation (DA)

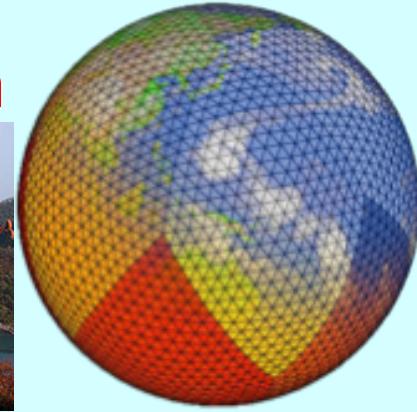
Observations



©Vaisala

Numerical models

Data Assimilation



Data assimilation best combines observations and a model, and brings synergy.

Data Assimilation (DA)

Observations



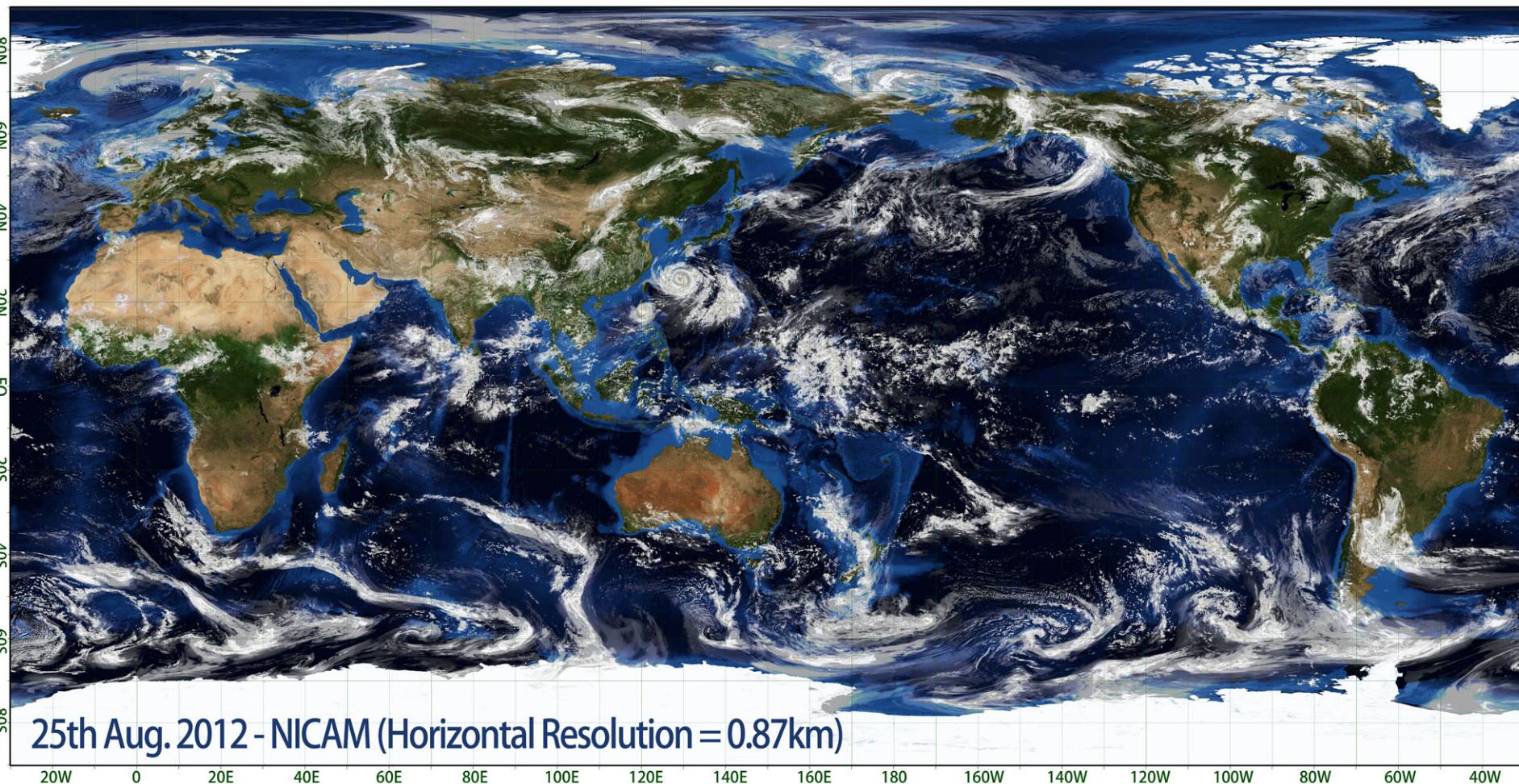
Numerical models

Data Assimilation



>2

Global 870-m simulation (*Miyamoto et al. 2013*)



©JAMSTEC・AORI (SPIRE Field3), RIKEN/
AICS
Visualized by Ryuji Yoshida

Computers getting more powerful...

- With the “post-K” supercomputer (~2020), we can afford **100 members** of global 870-m simulation.
- Two directions:
 - High resolution
 - Large ensemble



The Japanese 10-Peta-Flops K computer

©RIKEN

News & Media



Home > News & Media > Press Releases > Press Releases 2014 >

Press Release

[← Previous](#)[↑ Index](#)[→ Next](#)

July 23, 2014

 [Like](#)

59

 [Tweet](#)

21



K computer runs largest ever ensemble simulation of global weather

Ensemble forecasting, called ensembles, try to improve forecasts. The 10-petaFLOPS K computer has succeeded in using data assimilation to reduce the range of uncertainties.

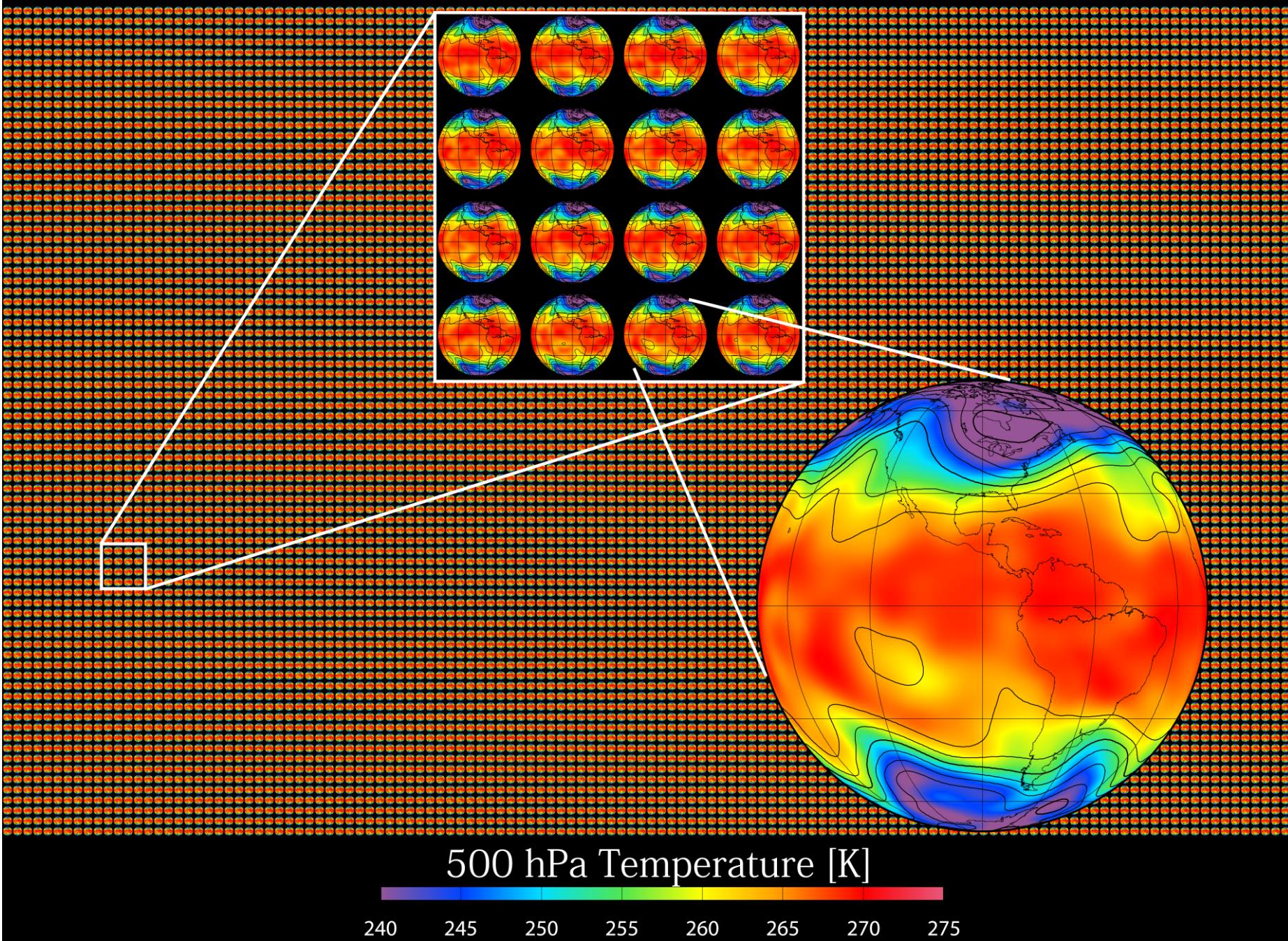
A simulated study using the T30/L7 SPEEDY AGCM

The assimilation of the 10,240 ensemble data sets was made possible by a cross-disciplinary collaboration of data assimilation experts and eigenvalue solver scientists at RIKEN AICS. The "Local Ensemble Transform Kalman Filter" (LETKF), an already efficient system, was further improved by a factor of eight using the "EigenExa" high-performance eigenvalue solver software, making possible a three-week computation of data from the 10,240 ensembles for simulated global weather. By analyzing the 10,240 equally probable estimates of atmospheric states, the team discovered that faraway observations, even going beyond 10,000 kilometers in distance, may have an immediate impact on eventual state of the estimation. This finding suggests the need for further research on advanced methods that can make better use of faraway observations, as this could potentially lead to an improvement of weather forecasts.

News & Media

[Press Center](#)[Press Releases](#)[2014](#)[2009](#)[2008](#)[2007](#)[2006](#)[2005](#)[News](#)[Events & Symposia](#)[Publications](#)[Videos](#)

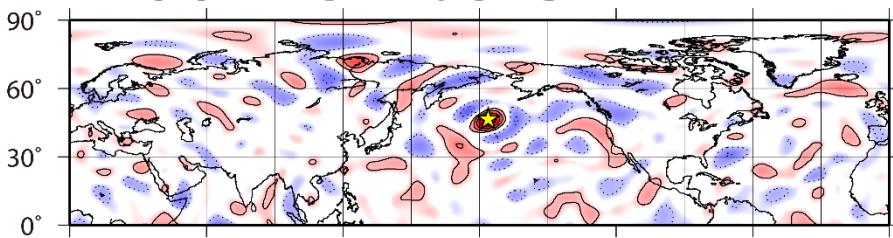
10240 parallel earths



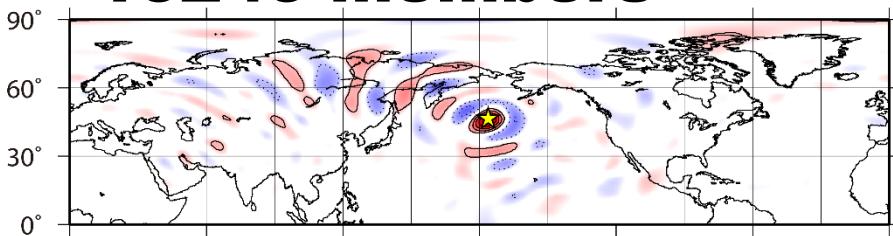
Advantage of large ensemble

(Miyoshi, Kondo, Imamura 2014)

100 members

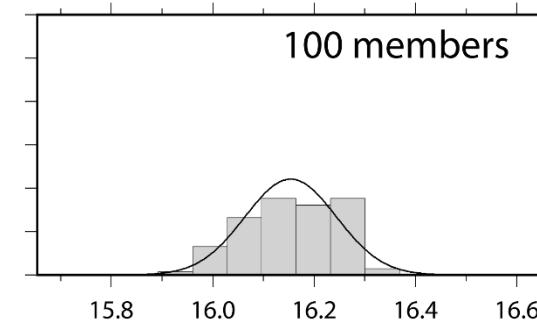


10240 members

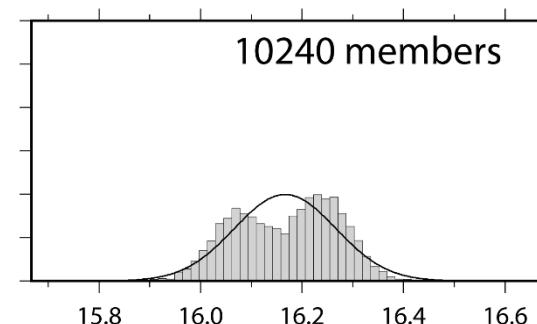


Sampling noise reduced

100 members



10240 members



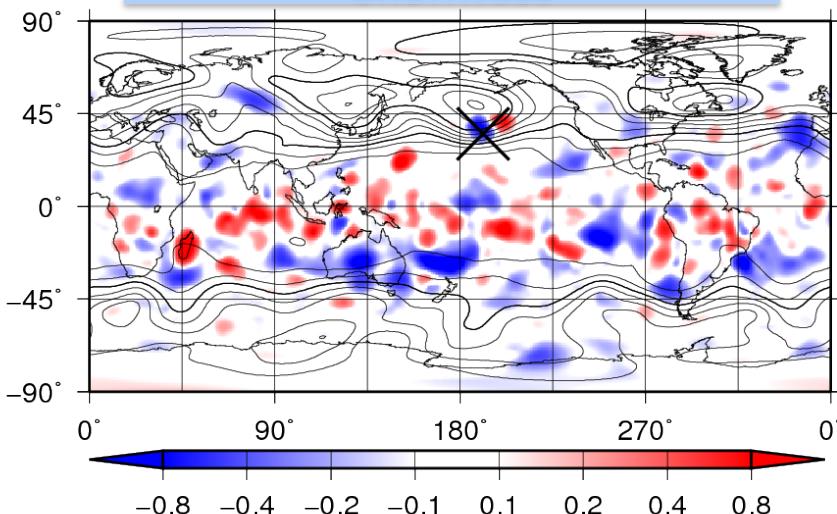
High-precision probabilistic representation

Non-Gaussianity

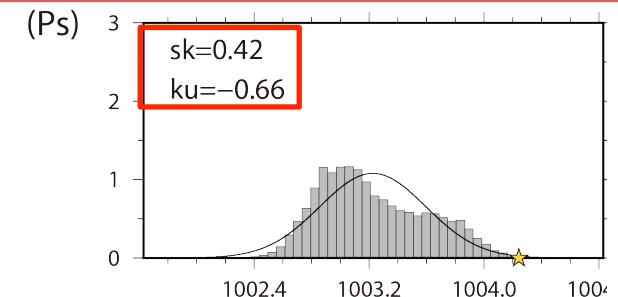
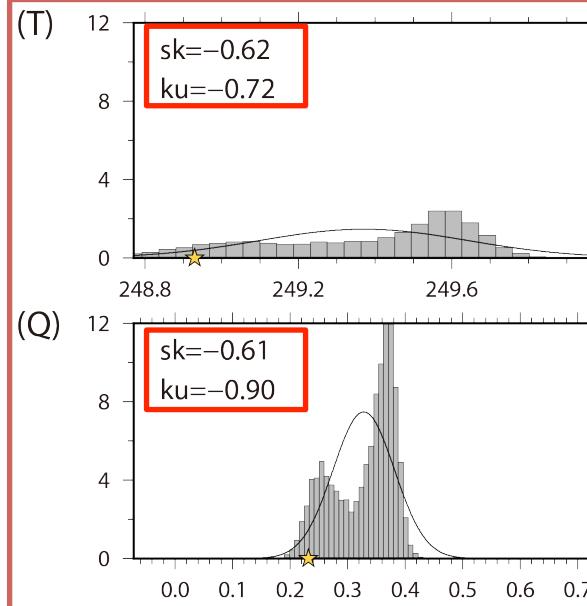
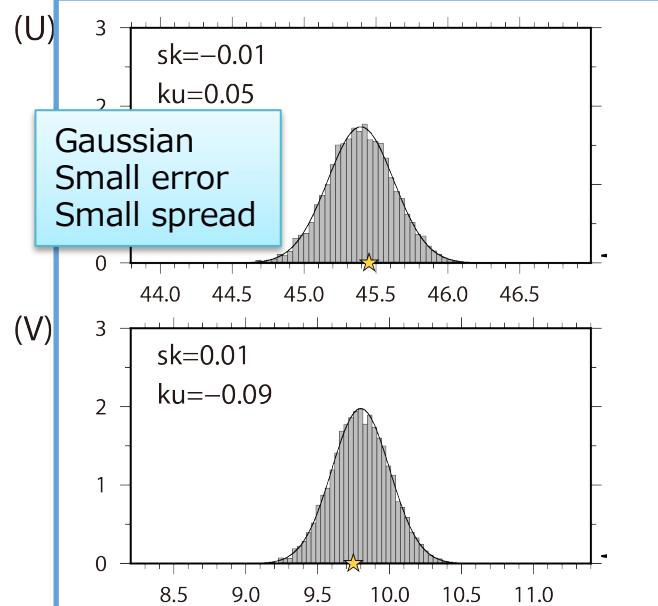
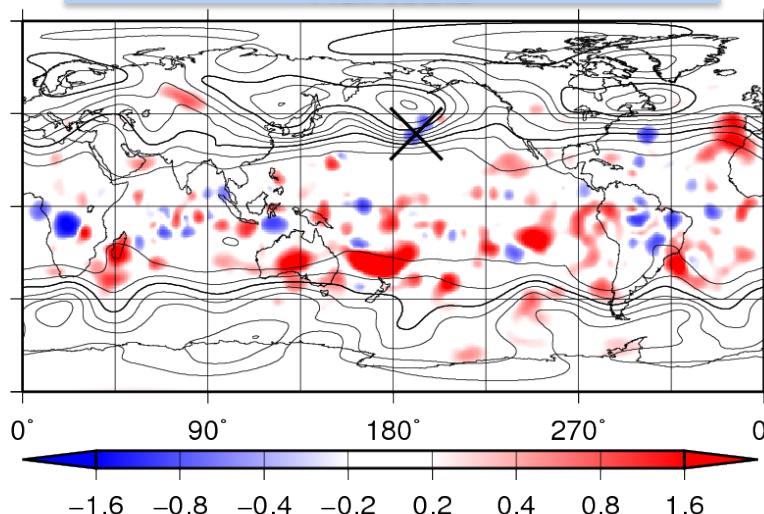
- **T at 4th level (00 UTC, 01/13/1982)**

(Kondo and Miyoshi 2015)

Skewness

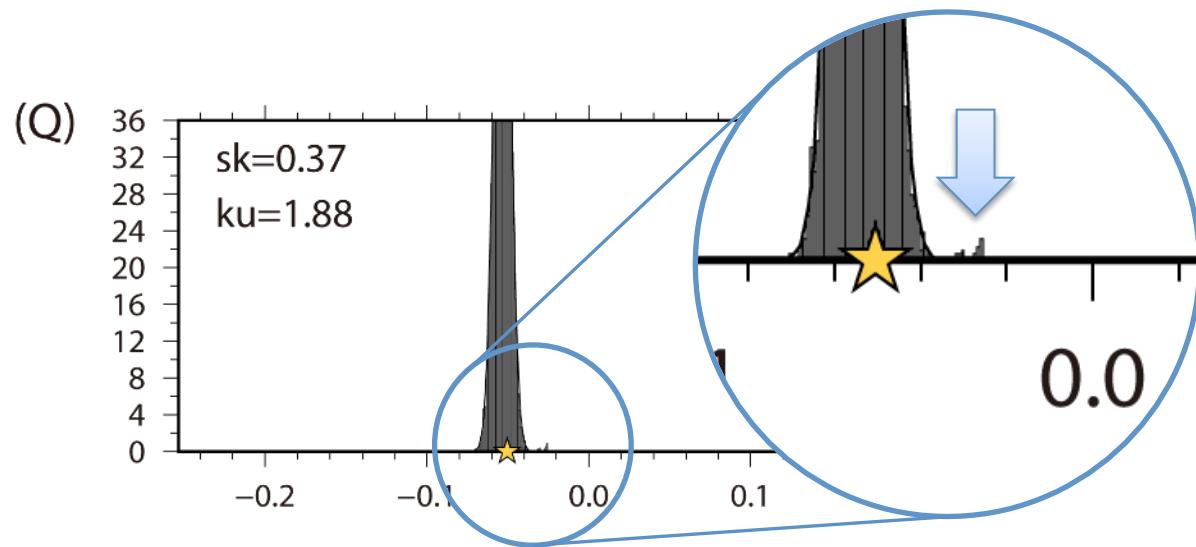


Kurtosis



Non-Gaussian
Large error
Large spread

Non-Gaussianity due to Outliers

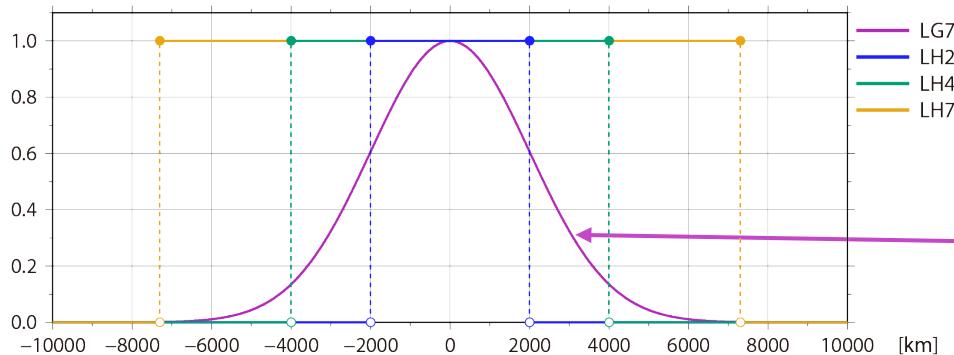


(Kondo and Miyoshi 2015)

Impact of far-away obs

Kondo&Miyoshi (2015)

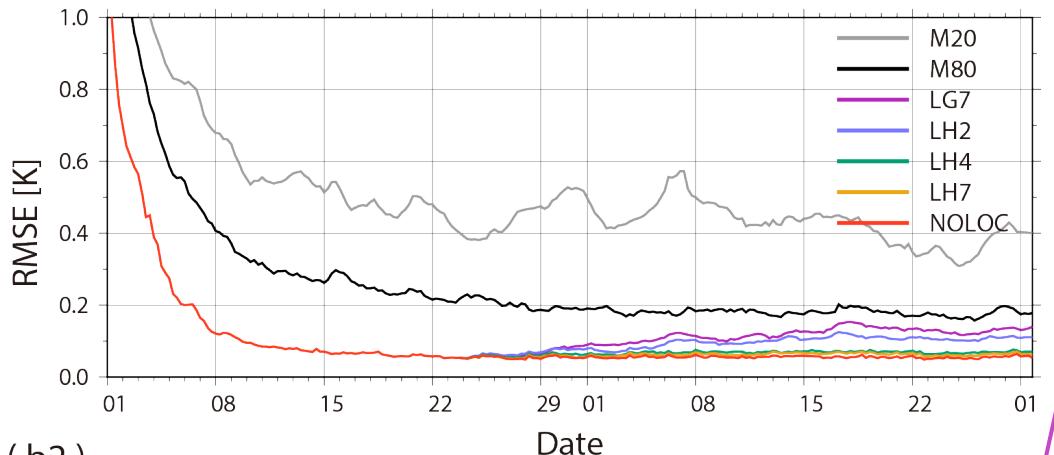
Localization weight



LHx: Heaviside step function
($\times 000$ km)

LG7: Gaussian tapering

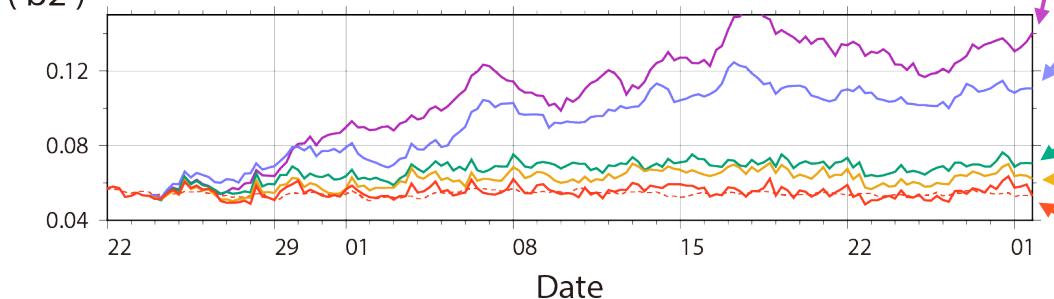
(b1) Analysis RMSEs (T , Lev = 2)



LG7: 7303 km Gaussian localization

LH2: 2000 km step localization

(b2)



LH4: 4000 km step localization

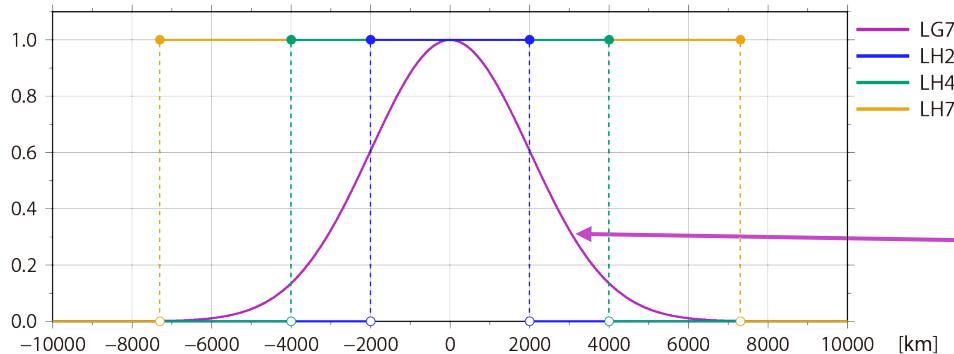
LH7: 7303 km step localization

NOLOC: no localization

Imbalance due to tapering

Kondo&Miyoshi (2015)

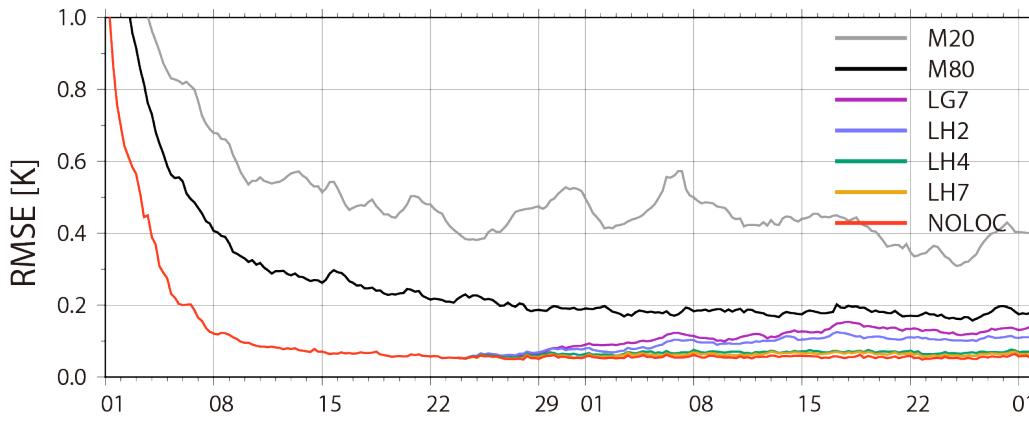
Localization weight



LG7: Gaussian tapering

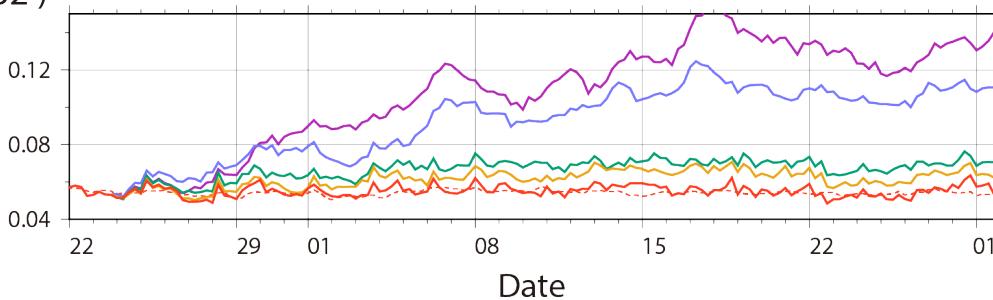
(b1)

Analysis RMSEs (T, Lev = 2)



(b2)

Date

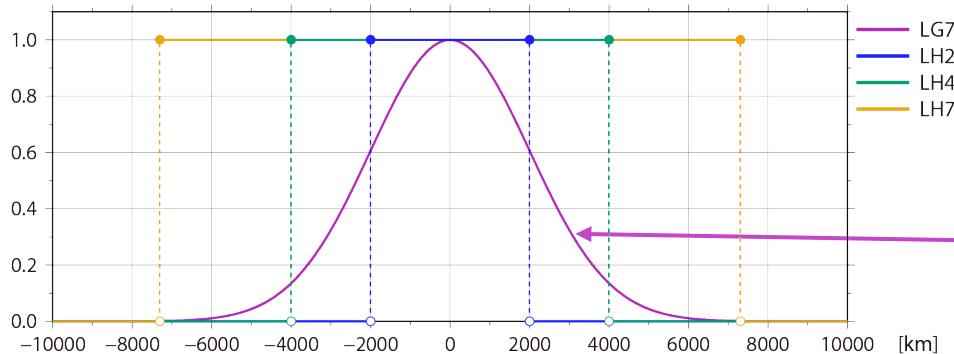


Unexpected. Why???

Imbalance due to tapering

Kondo&Miyoshi (2015)

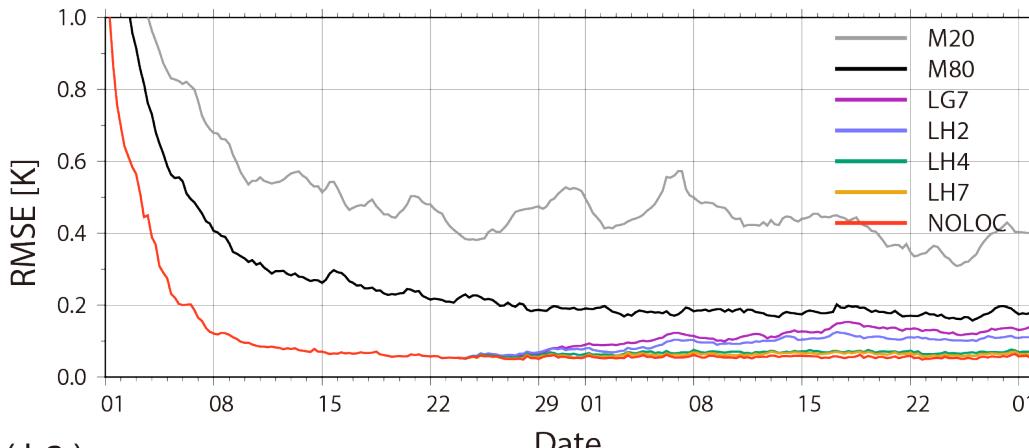
Localization weight



LG7: Gaussian tapering

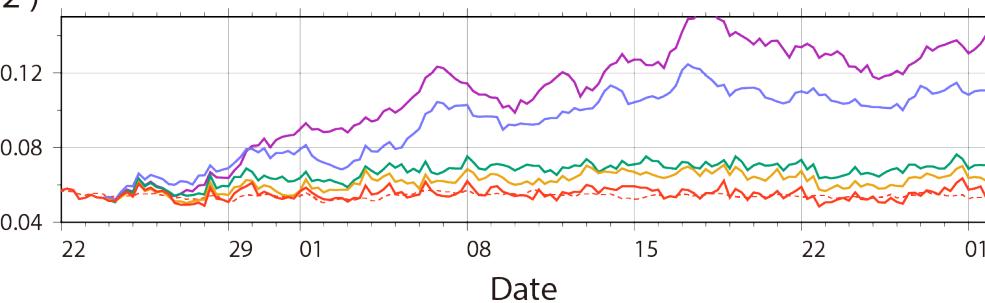
(b1)

Analysis RMSEs (T , Lev = 2)



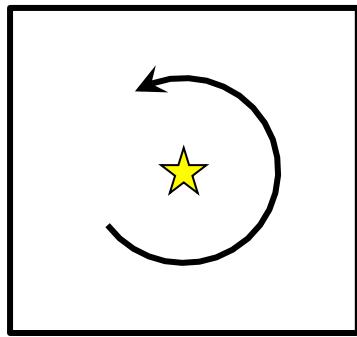
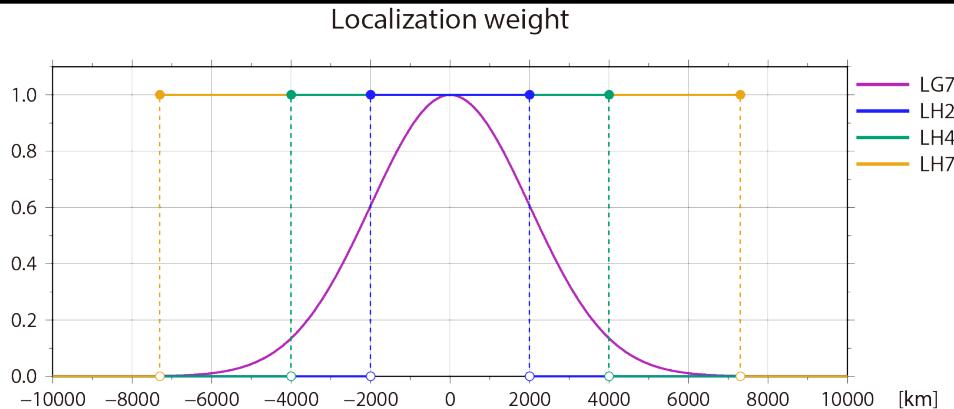
(b2)

likely due to the
imbalance by
localization tapering



Imbalance due to tapering

Kondo&Miyoshi (2015)



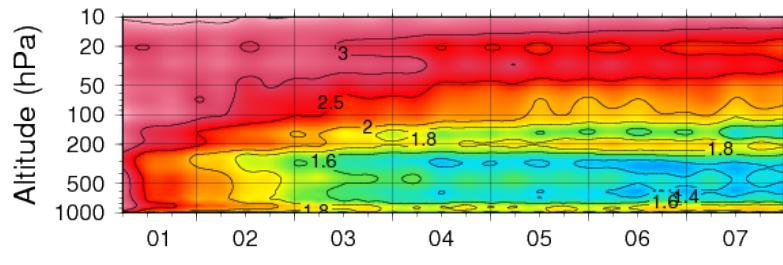
An observation of mass variable induces “balanced” rotating-flow increments.

Due to localization:
mass gradient increased,
flow speed reduced
→ Imbalance!

Simulated case → Real case (*ongoing*)

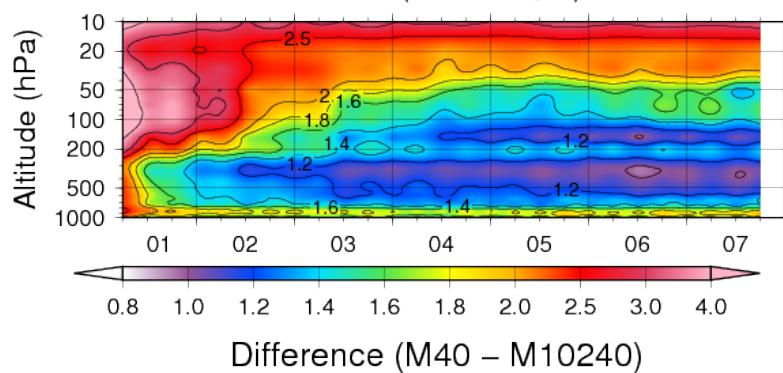
Temperature [K] RMS Diff (vs. ERA-Interim)

RMSD (M40, T)



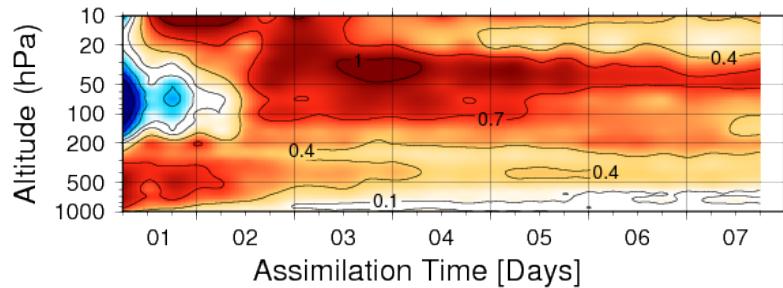
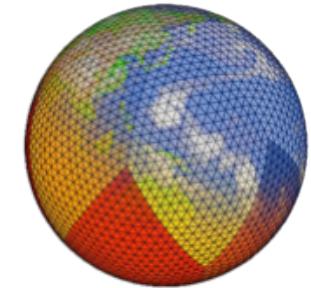
40 members

Kondo&Miyoshi
(2015)



10240 members

NICAM-LETKF
(Terasaki et al. 2015)



Improvements

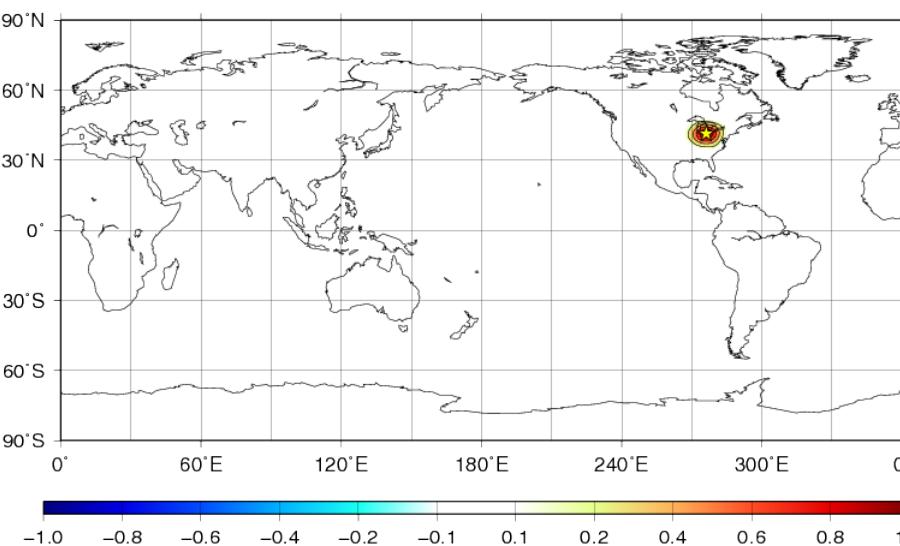
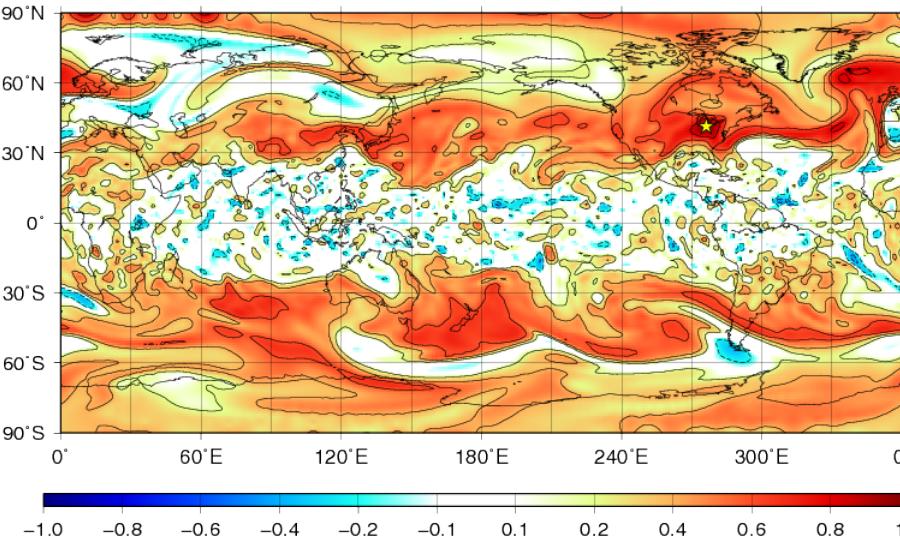
Degraded ←

→ Improved

Correlation patterns (Q at \sim 100 hPa)

40 members

Kondo, Miyoshi (2015)



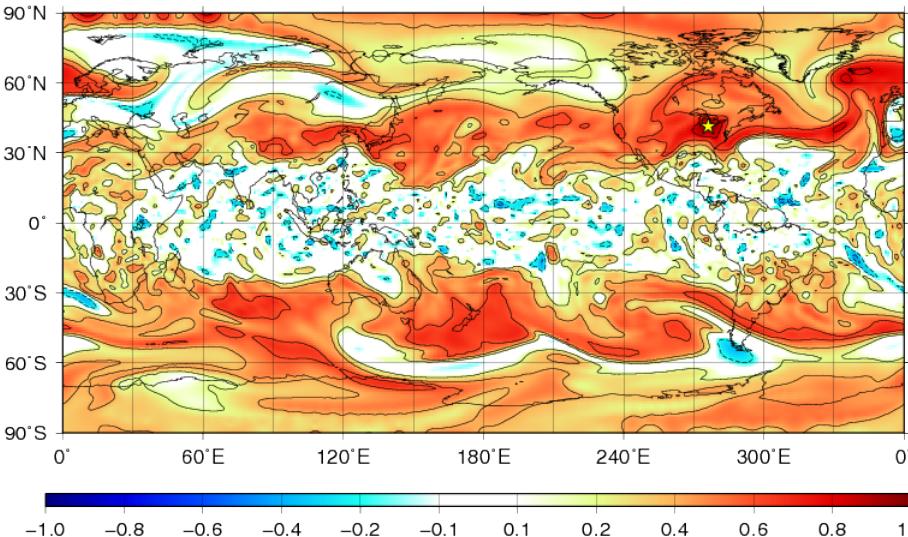
**Localized
($\sigma=400\text{km}$)**

This is what we use for EnKF
with 40 members.

11/8 00UTC after a week cycling

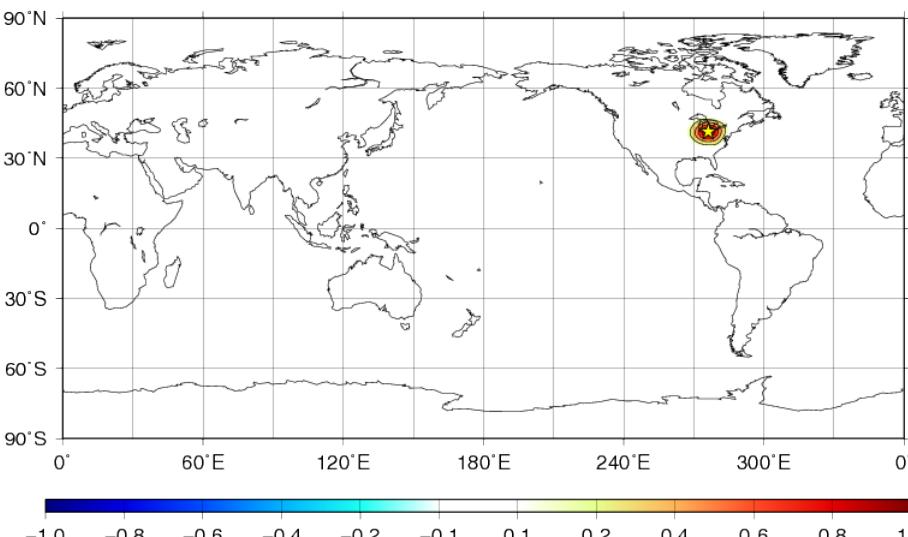
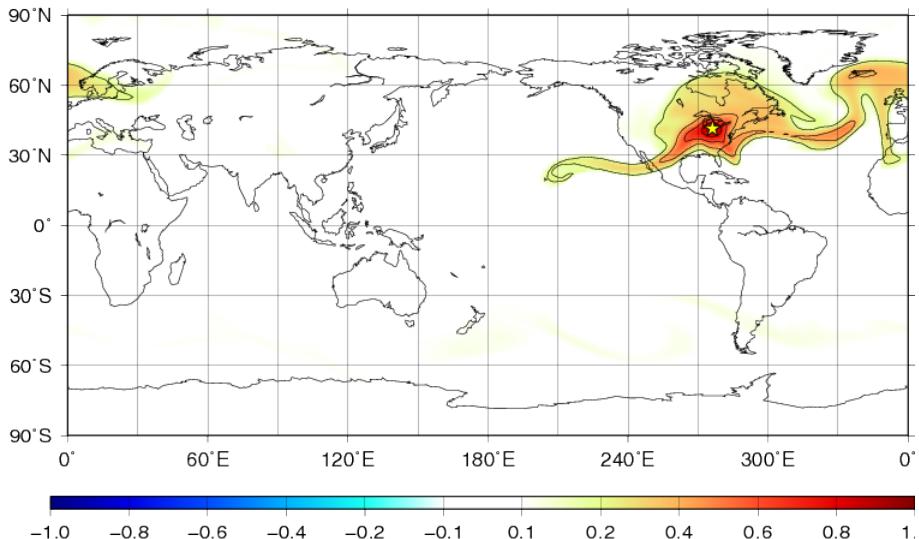
Correlation patterns (Q at \sim 100 hPa)

40 members



Kondo, Miyoshi (2015)

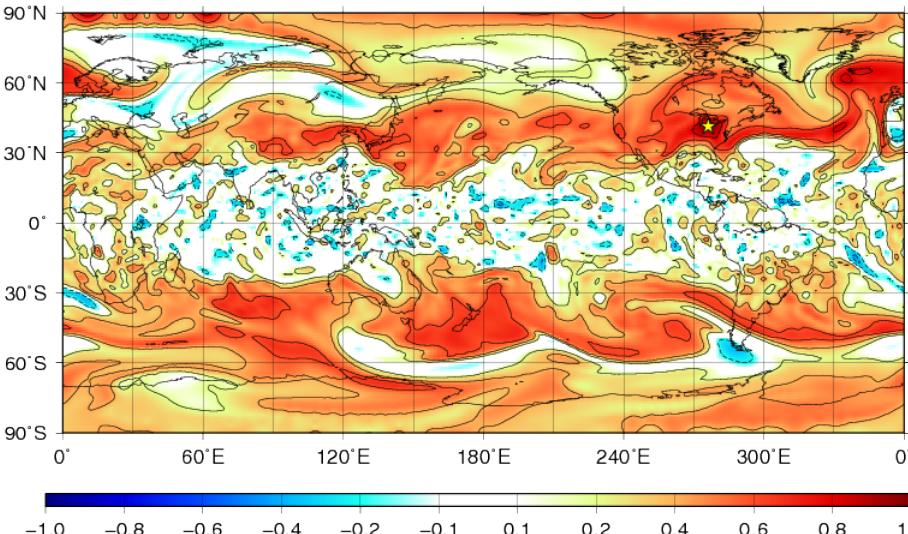
10240 members



11/8 00UTC after a week cycling

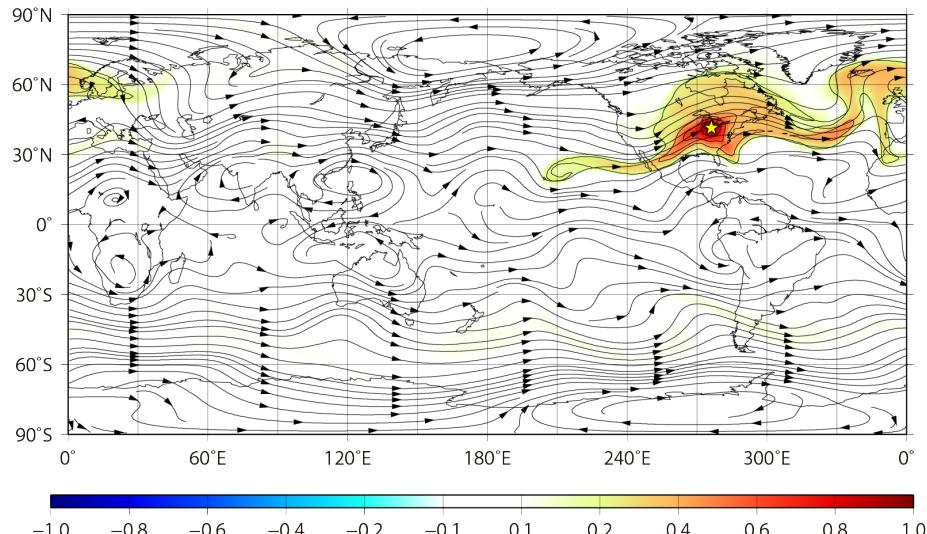
Correlation patterns (Q at \sim 100 hPa)

40 members

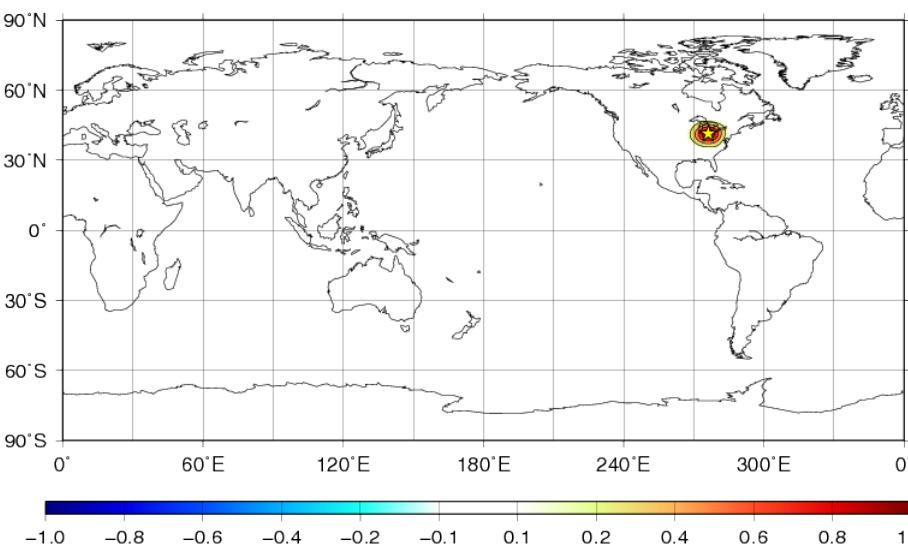


Kondo, Miyoshi (2015)

10240 members



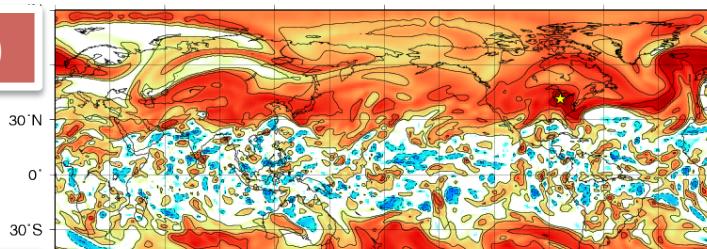
FLOW-DEPENDENT



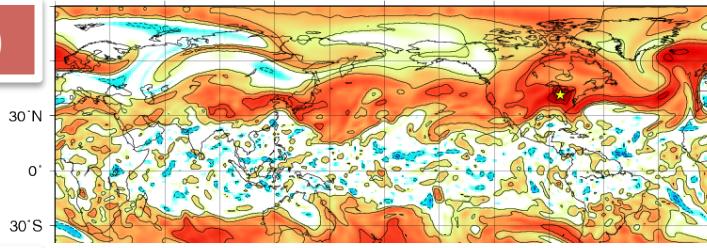
11/8 00UTC after a week cycling

With subsets of 10240 samples *Kondo&Miyoshi*

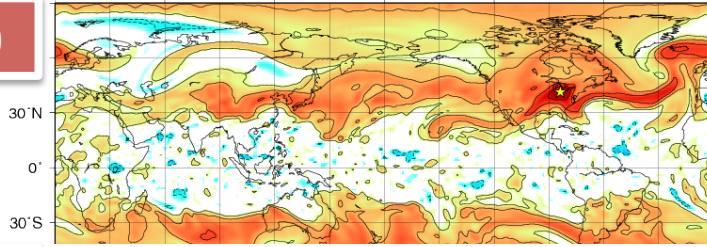
20



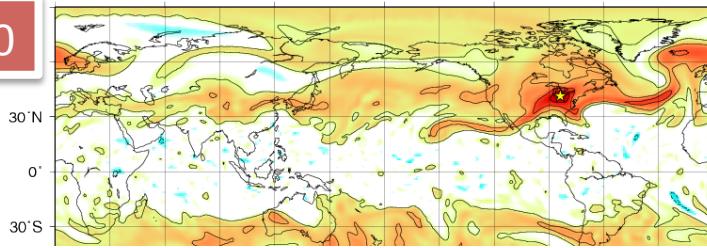
40



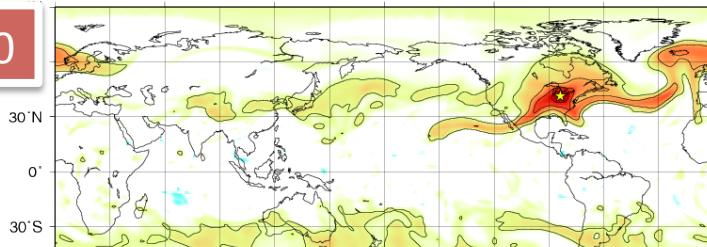
80



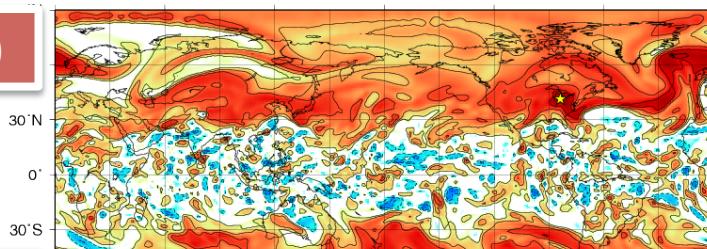
160



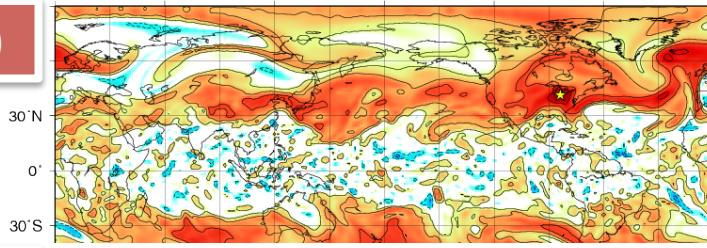
320



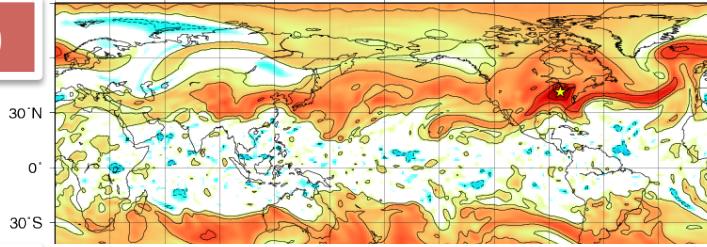
640



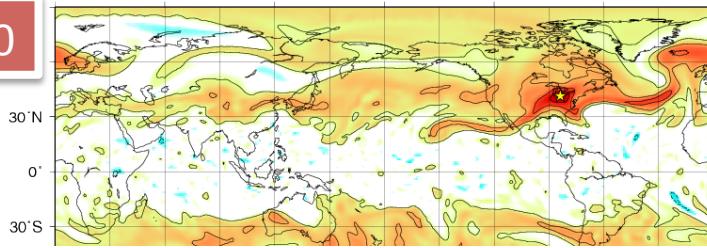
1280



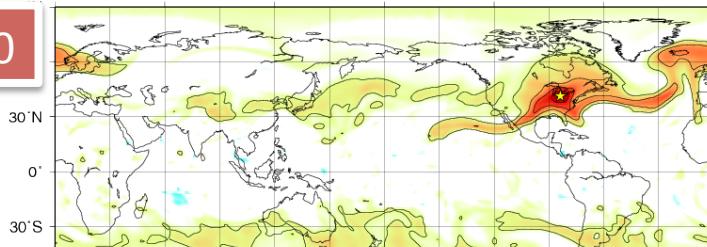
2560



5120



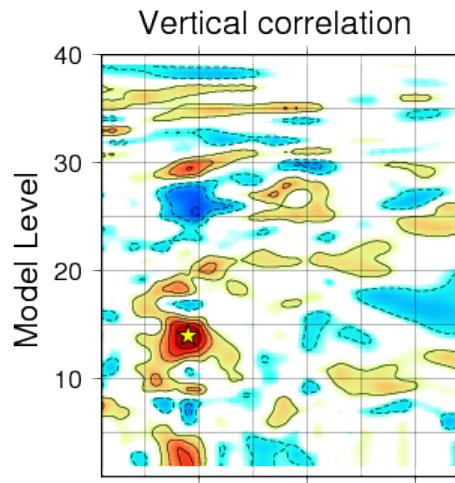
10240



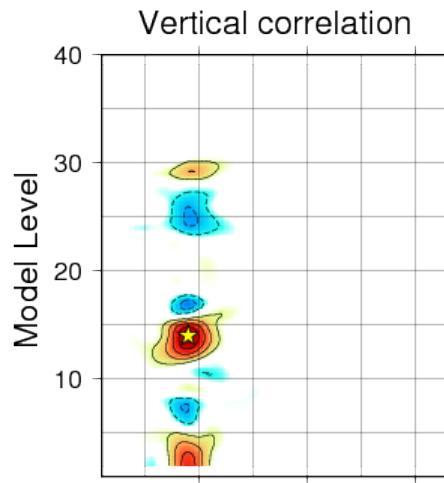
(2015)

Explore vertical correl. with big ensemble

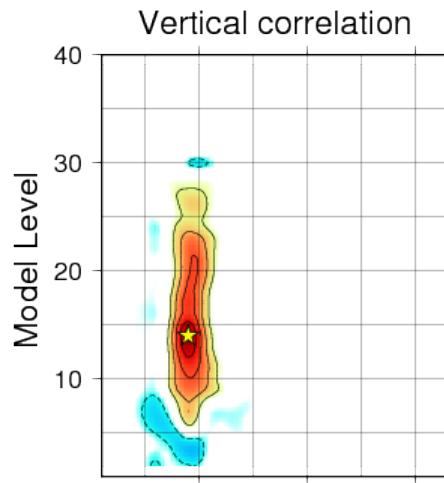
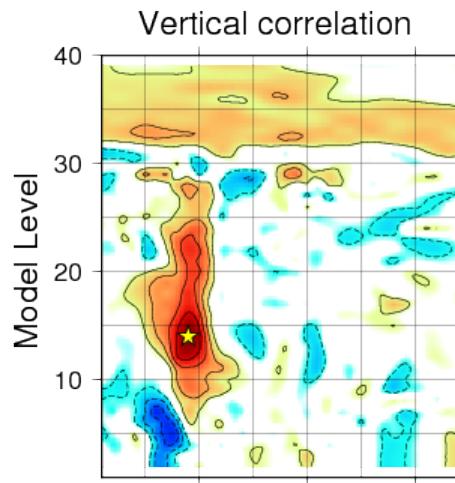
40 members



10240 members

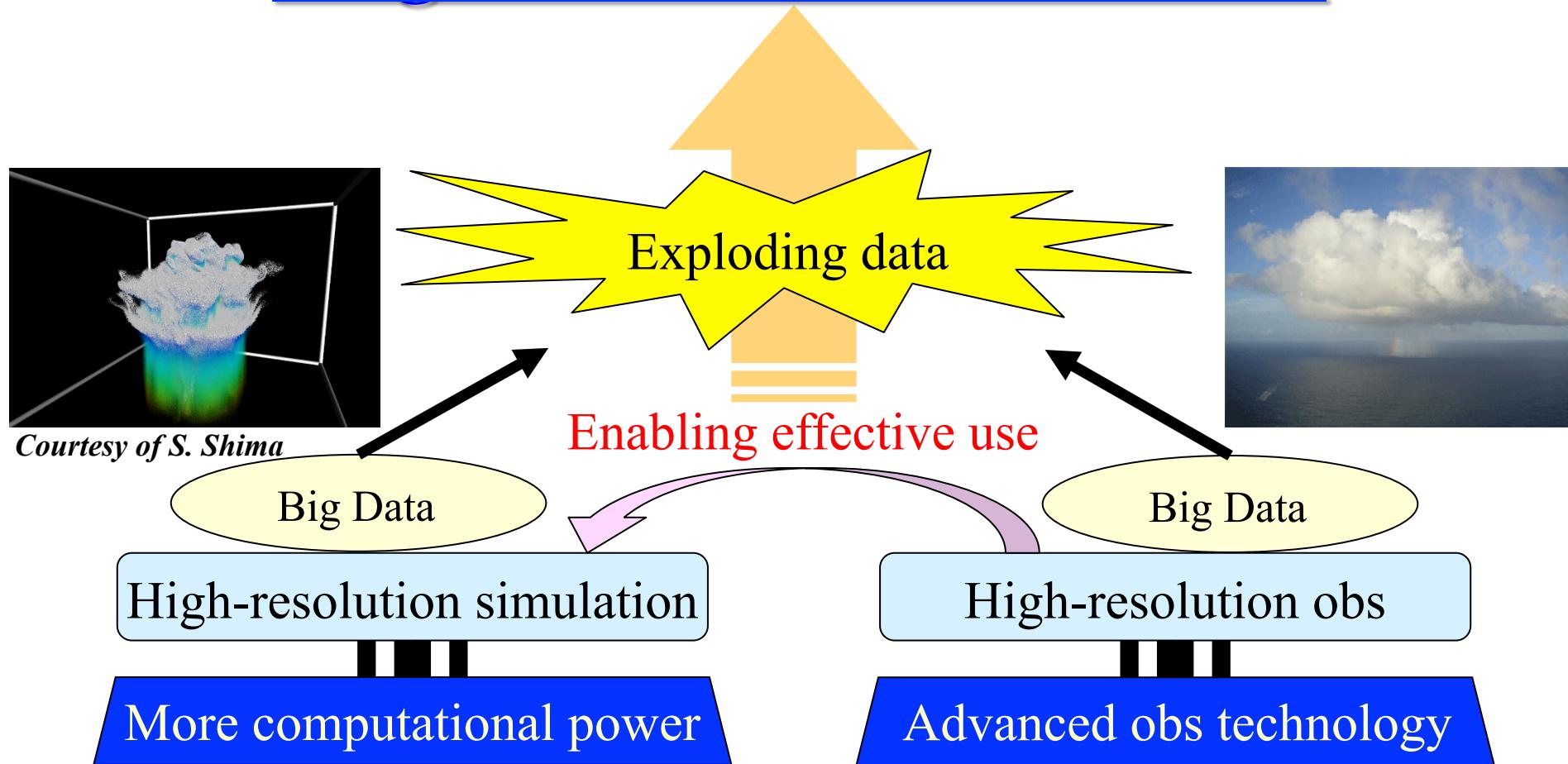


Humidity



Toward next 20 years of DA

“Big Data Assimilation” Era



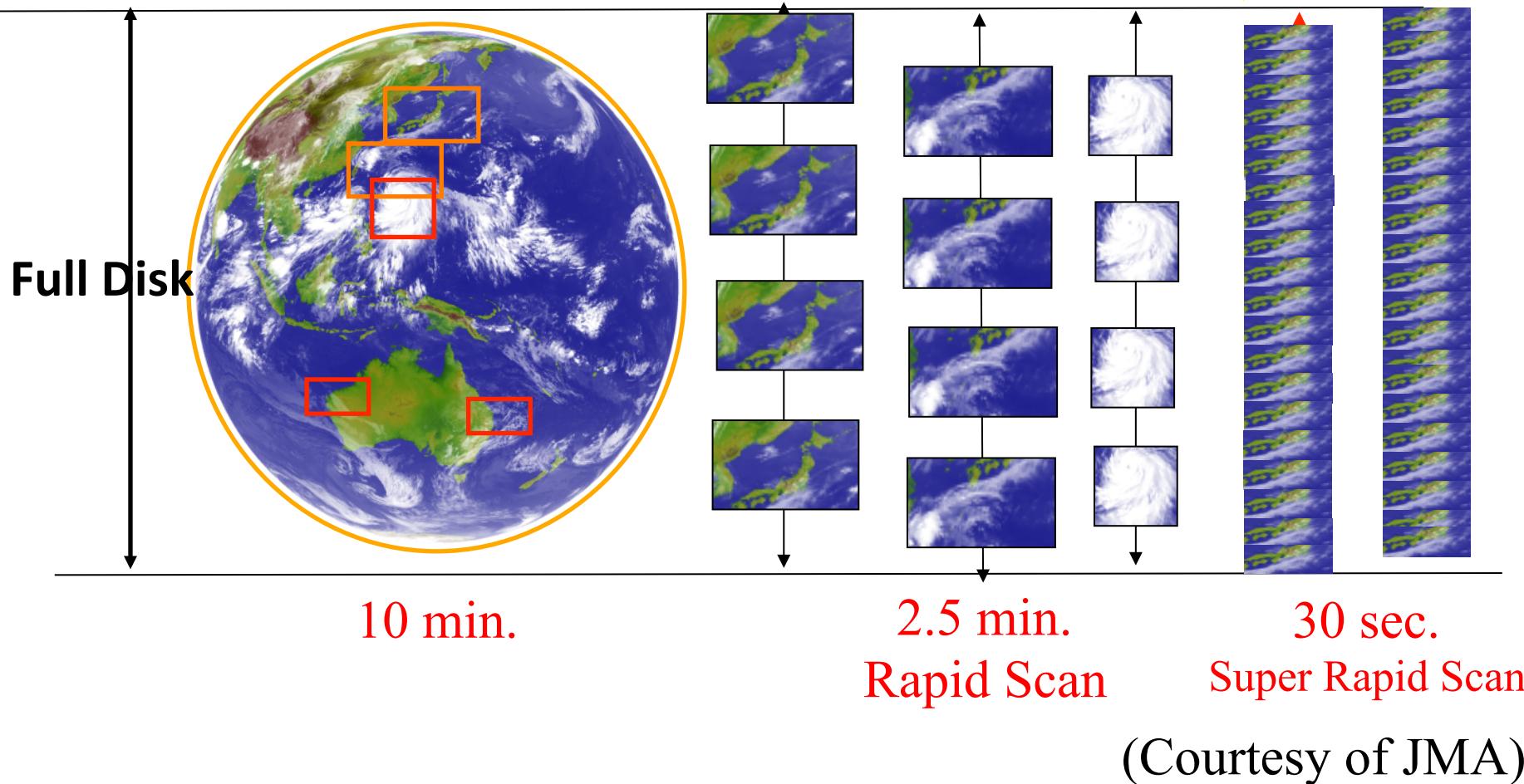
Next-generation geostationary satellite

Himawari-8 was launched successfully on 7 October 2014.

Himawari-9 will be launched in 2016.

Full operations started on 7/7/2015!!

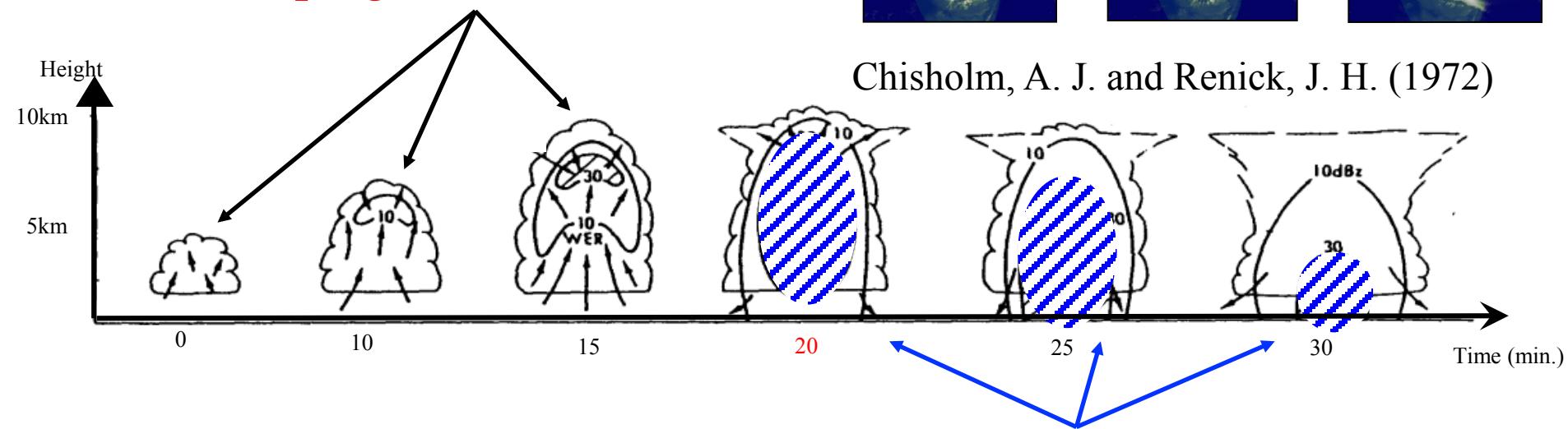
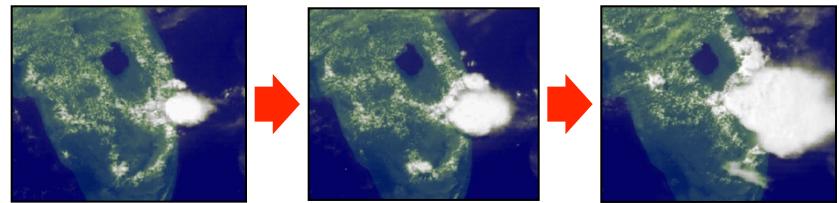
*Super Rapid Scan
every 30 seconds*



Rapid scan effective for convections

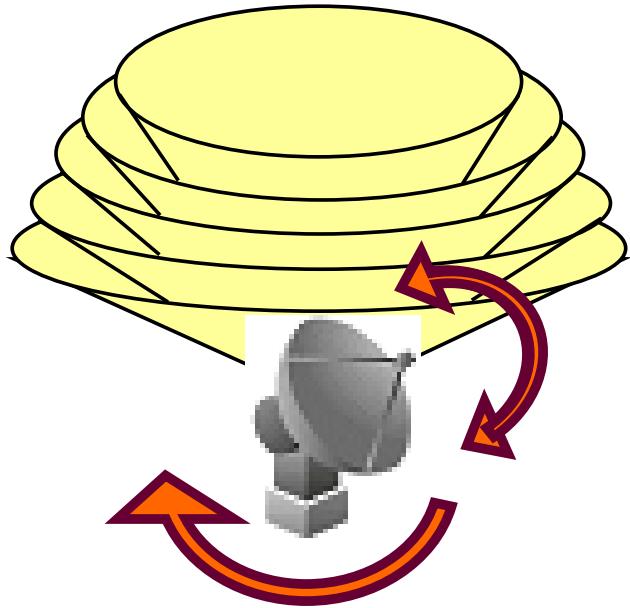
Typical lifetime of a convective system ~ 30 min.

Satellite imagery captures developing convections.



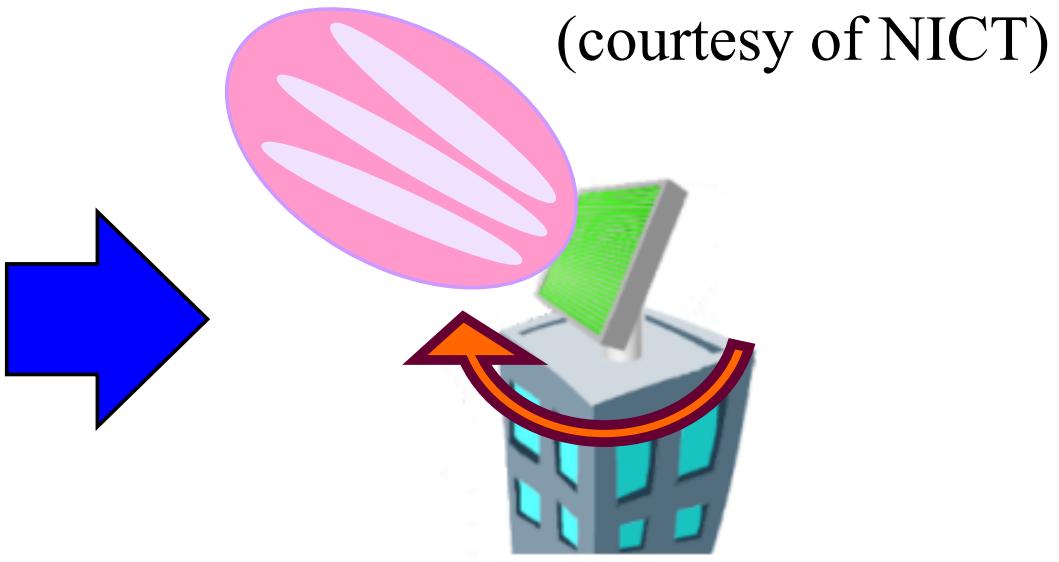
Radar captures rain particles
after the developing stage.
(may be too late...)

Phased Array Radar



Conventional Radar

~15 scan angles
Every 5-10 minutes

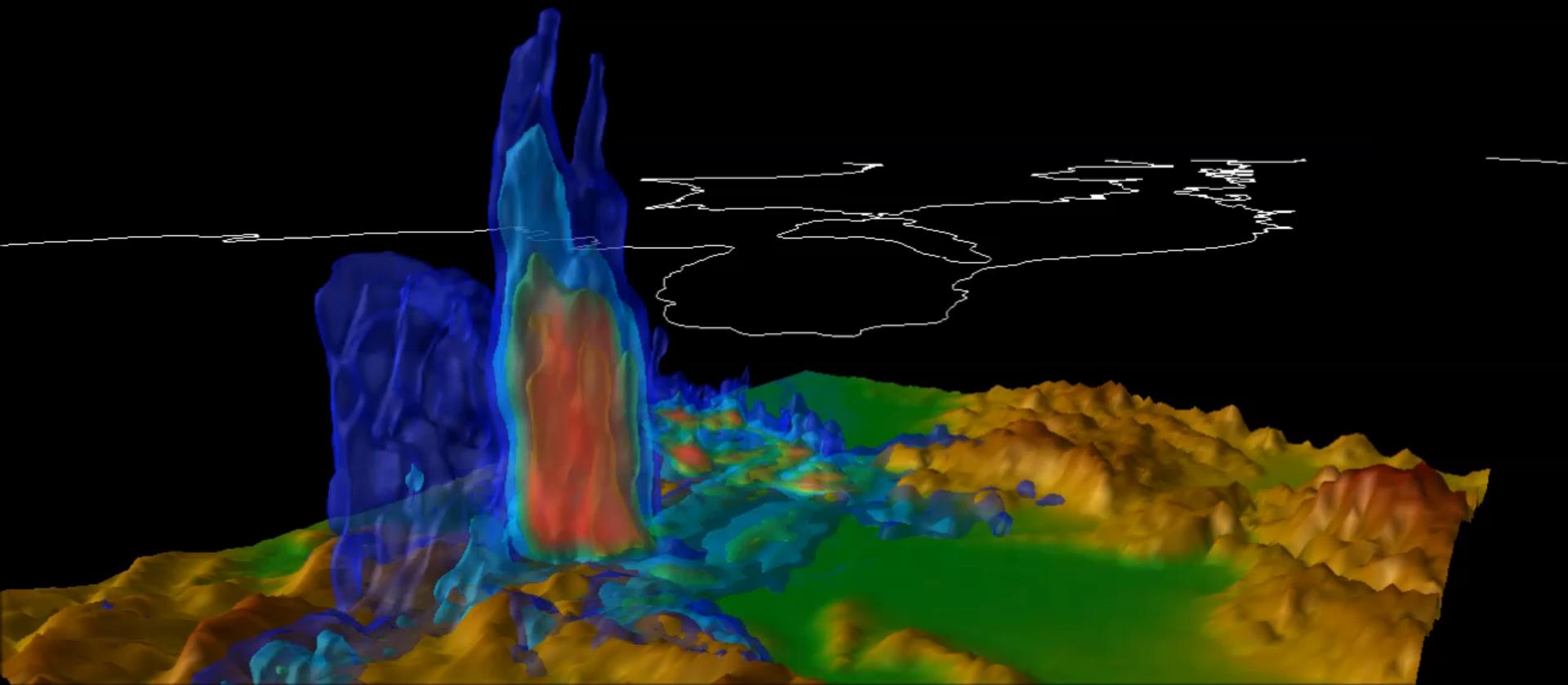


Phased Array Radar

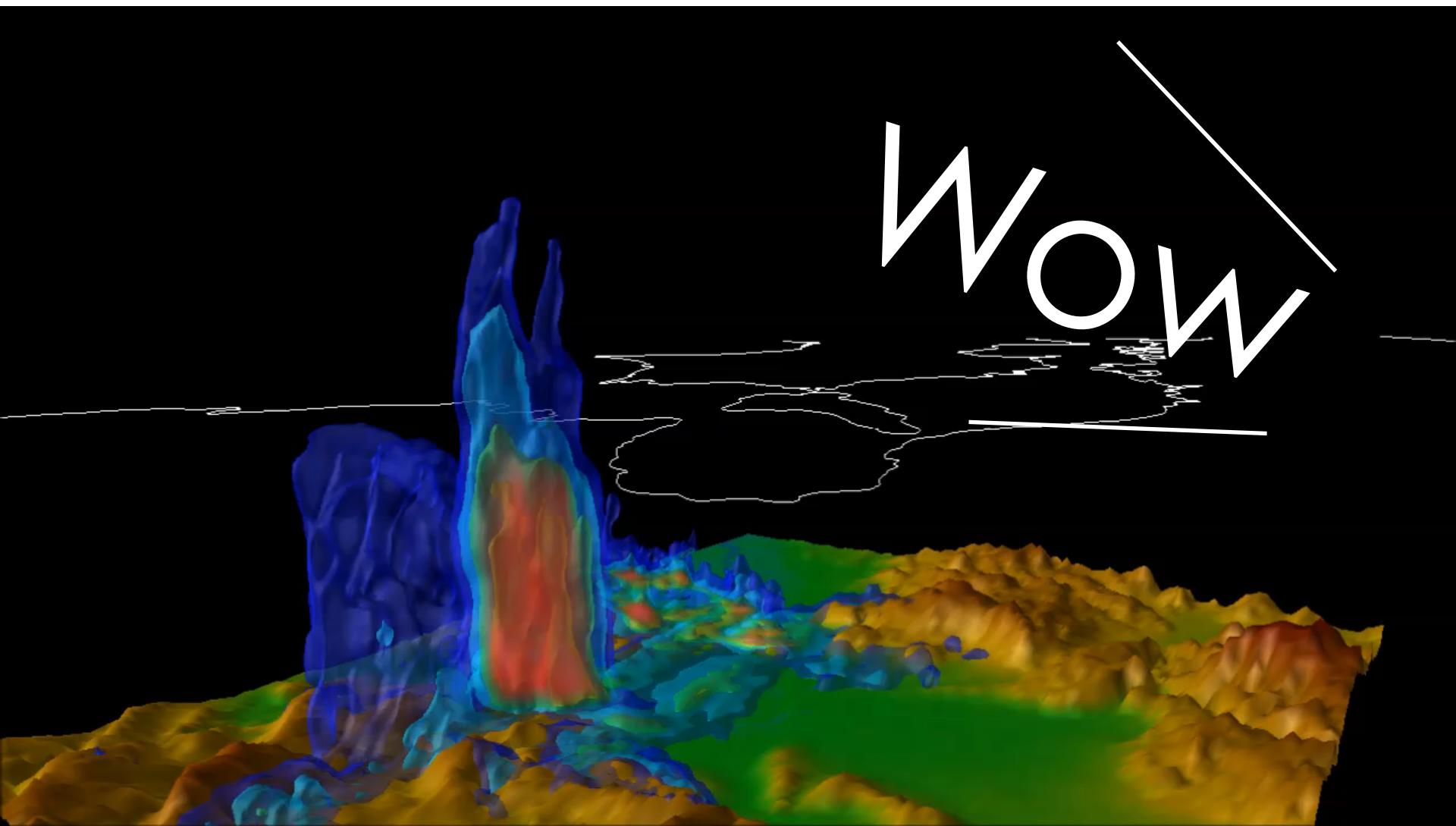
~100 scan angles
Every 10-30 seconds

(courtesy of NICT)

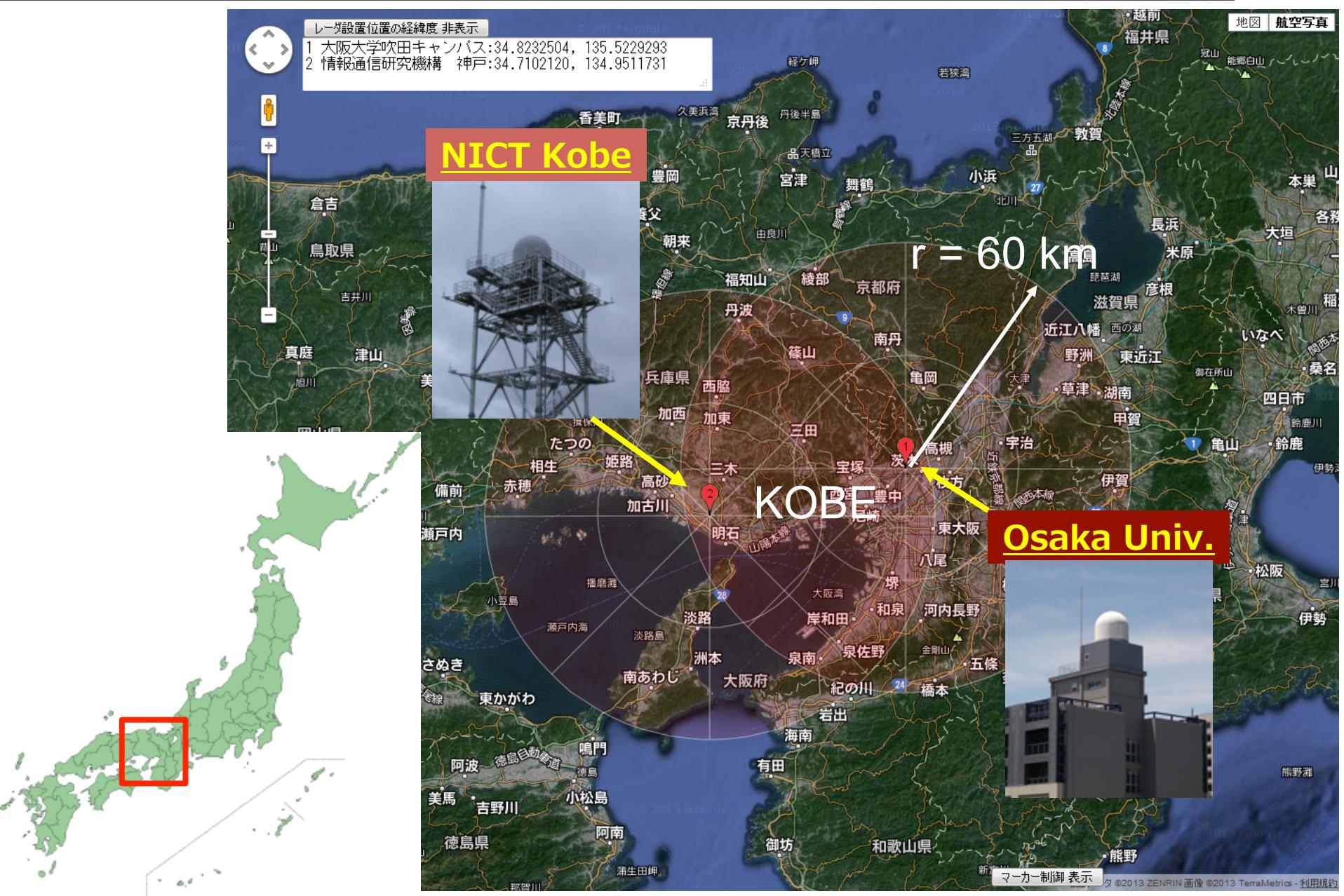
Conventional Radar (every 5 min.)



Phased Array Radar (every 30 sec.)



Two PAR in Kobe area



Exploring new data: live-camera images?



1. Reduced/extracted information (e.g., weather type, visibility)
(challenge) Automated image processing

2. Simulating images from model outputs
(challenge) precise 3-dimensional radiation model

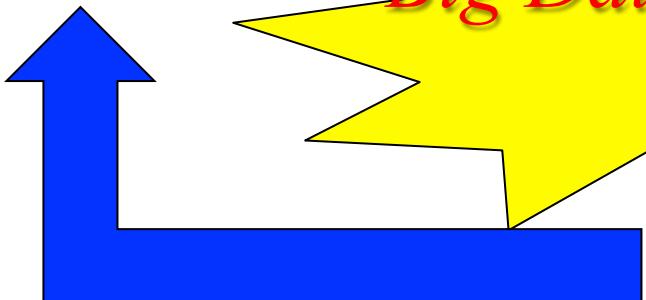
Towards “Big Data Assimilation”

High-resolution simulation

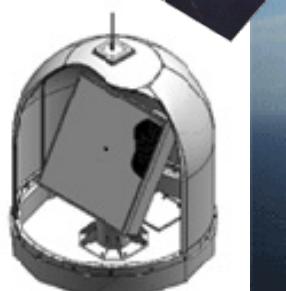


Combination of
next-generation technologies

“Big Data Assimilation”



Improving simulations



High-resolution observation

Storm forecasting with Big Data Assimilation

●親水公園で水遊び



水位は 10分間で約 1m30cm も上昇

5 people died in Kobe on July 28, 2008,
due to local heavy rainfall



Japan Meteorological Agency



増水直前

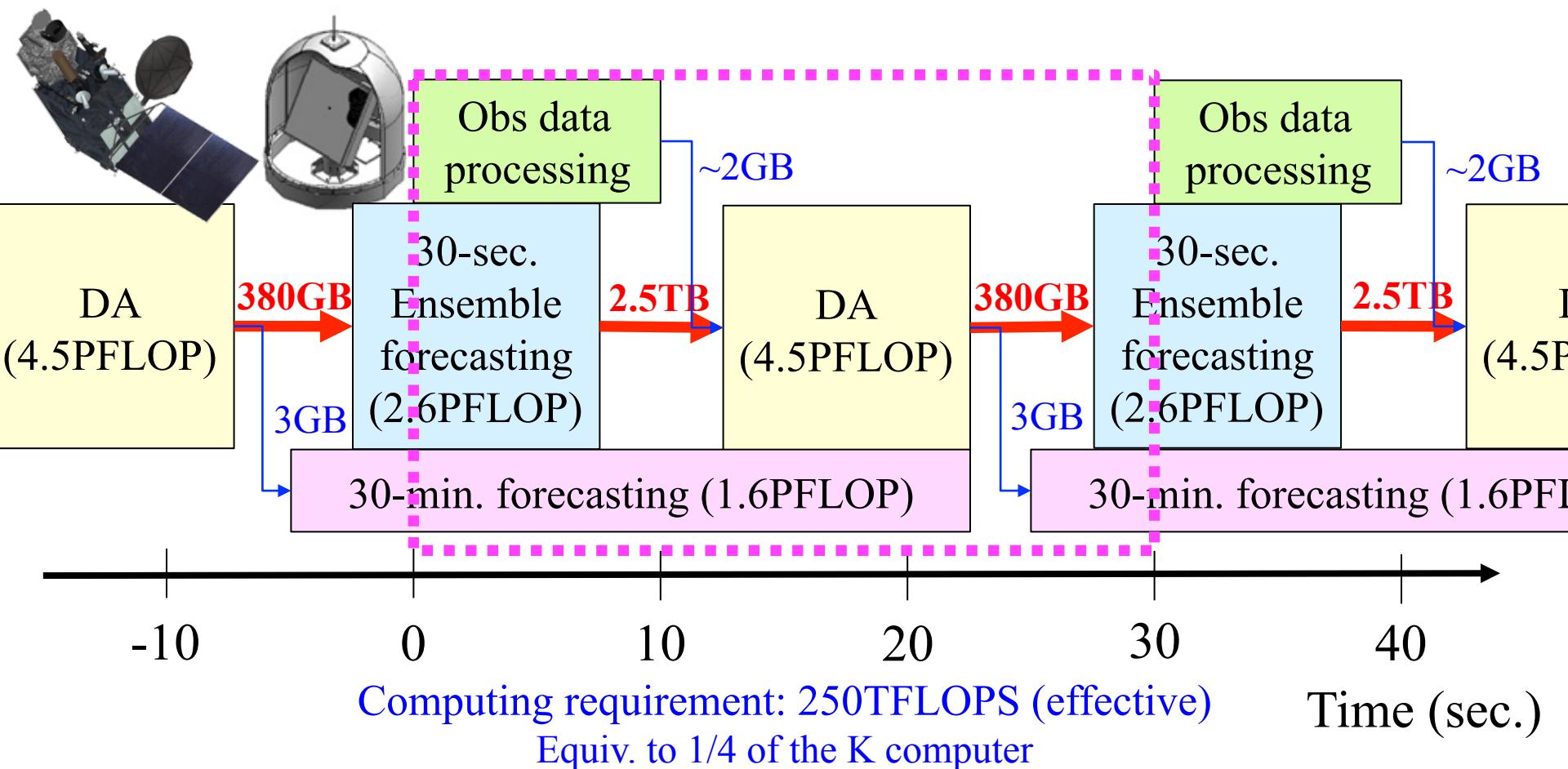
Only in 10 min.



増水時

Goal: 30-min forecasting of local severe weather through
Big Data Assimilation innovations.

Revolutionary super-rapid 30-sec. cycle

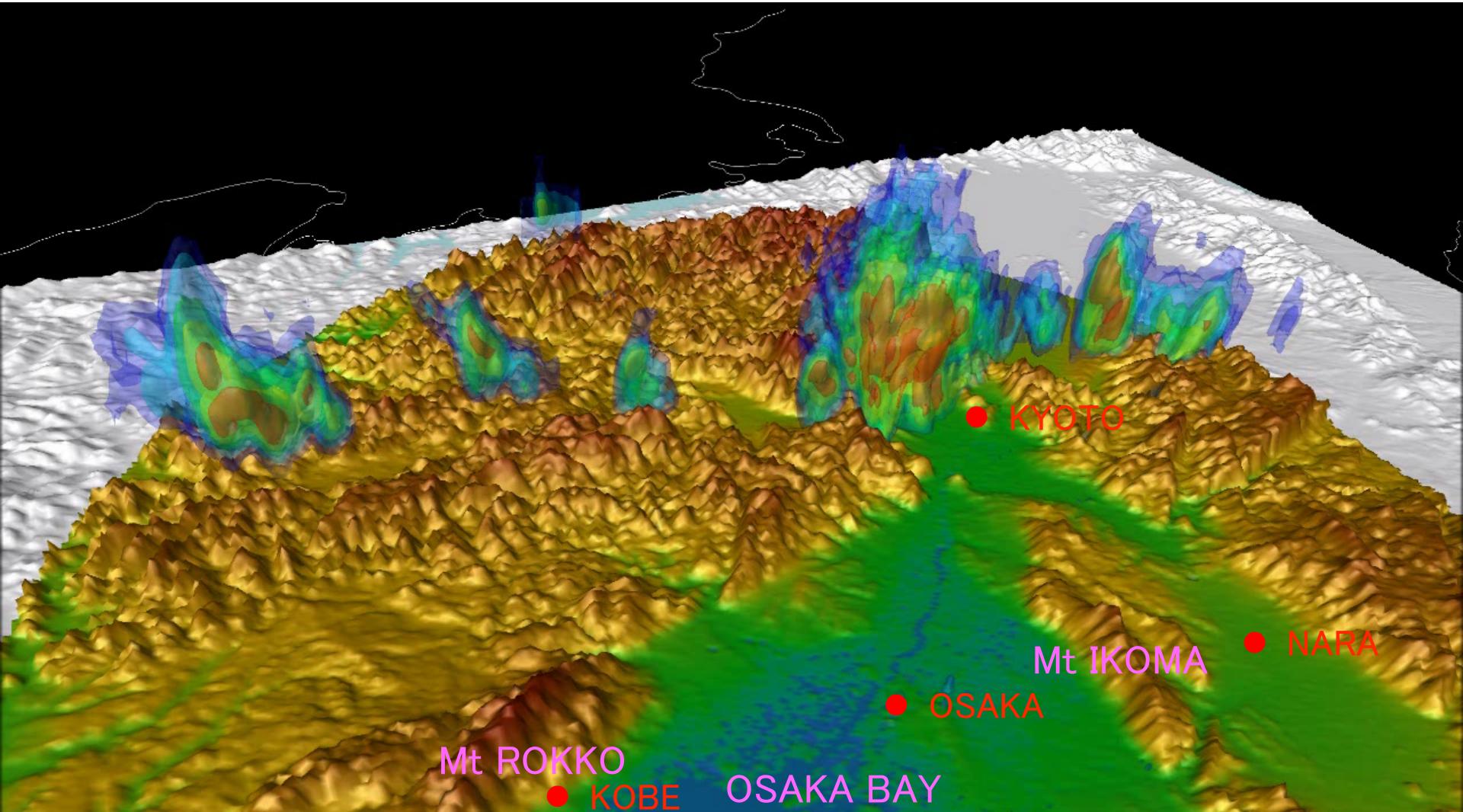


120 times more rapid than
hourly update cycles

A case selected for the first offline study

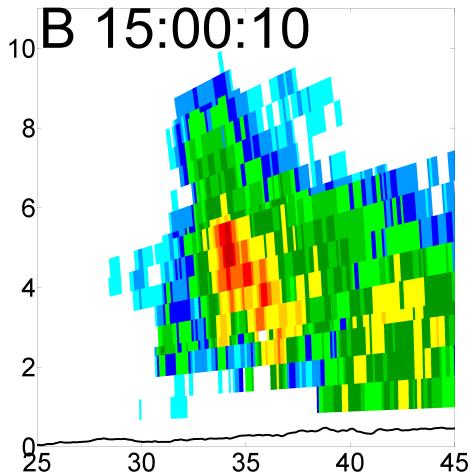
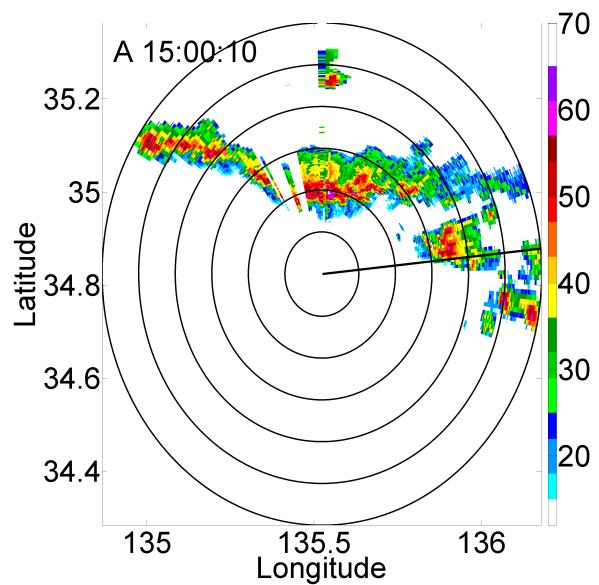


The top page of Yomiuri newspaper
on 14 July, 2013



10fps → 300x

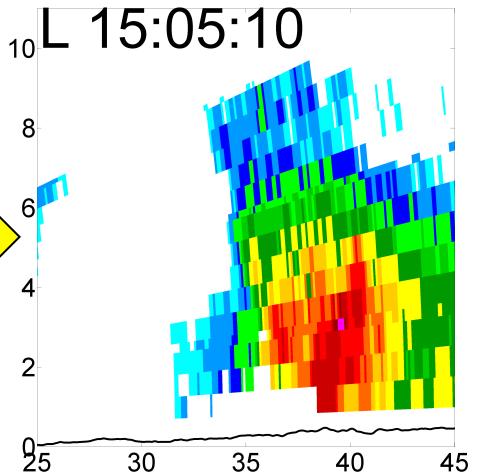
30-sec. and 5-min. evolutions



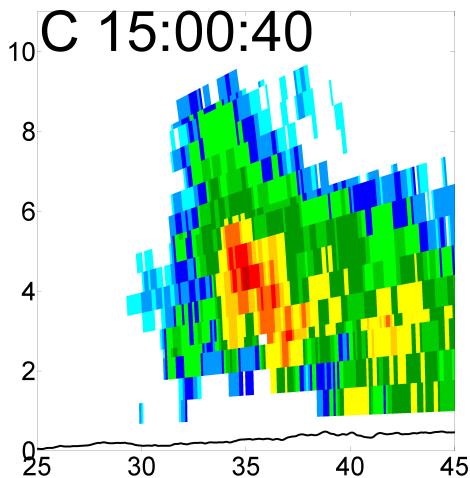
30 sec

(Miyoshi et al. 2015)

5 min



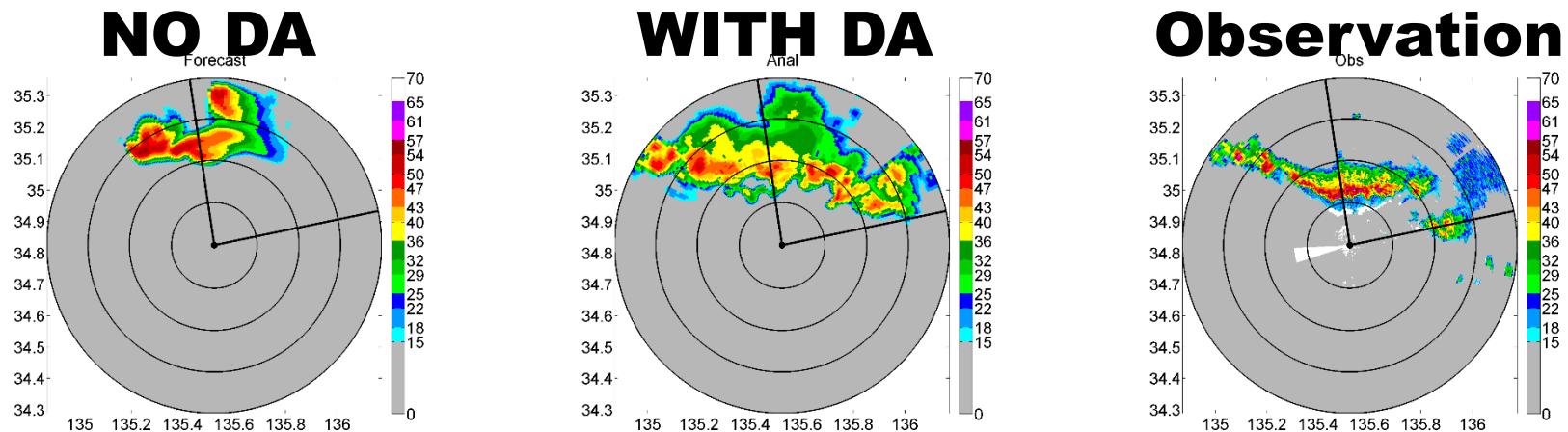
nonlinear evolution



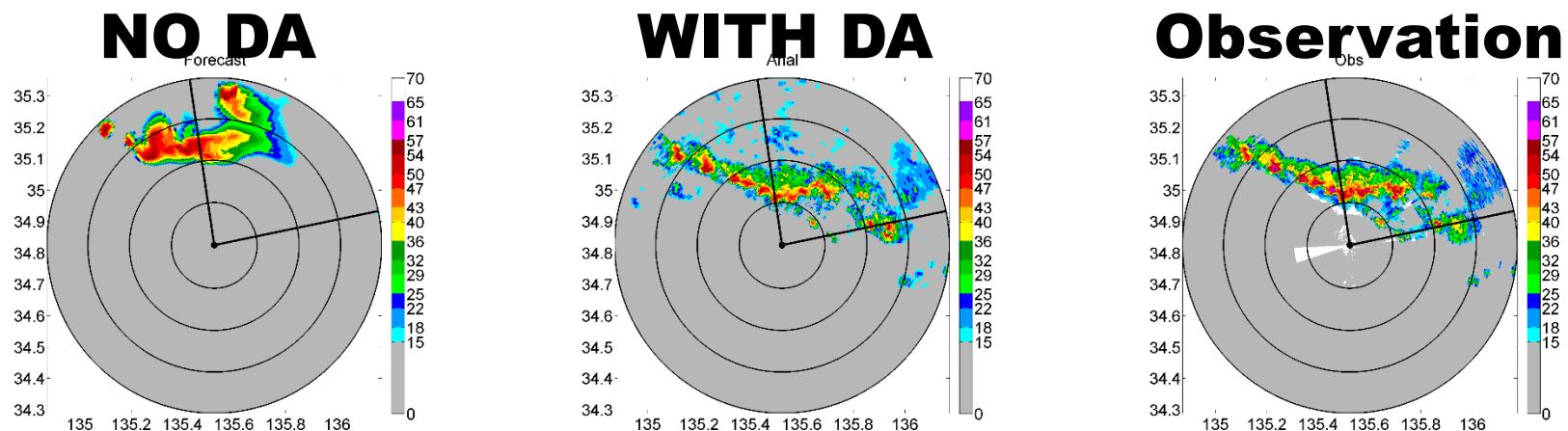
linear
evolution

Impact of “Big Data Assimilation”

15:00:30 JST after the first assimilation



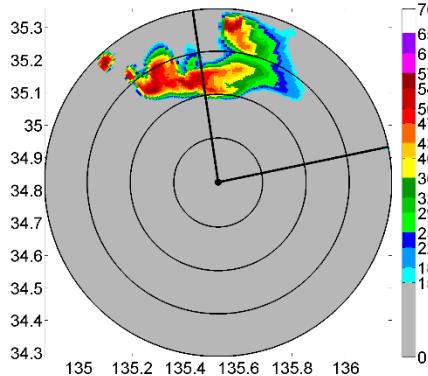
15:06:00 JST after the 12th assimilation



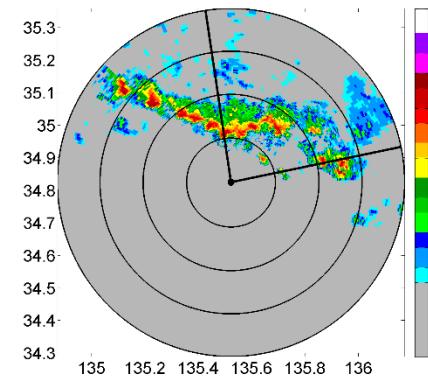
Vertical section

15:06:00 JST

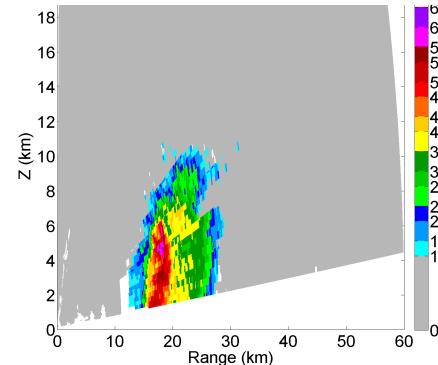
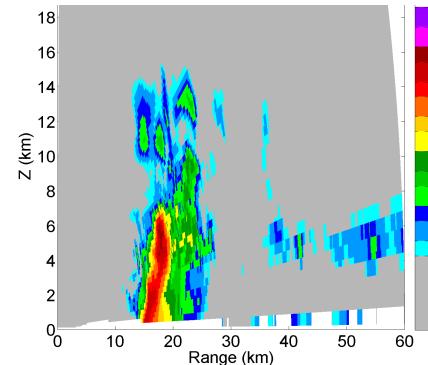
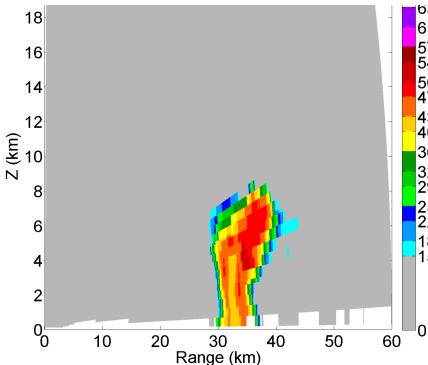
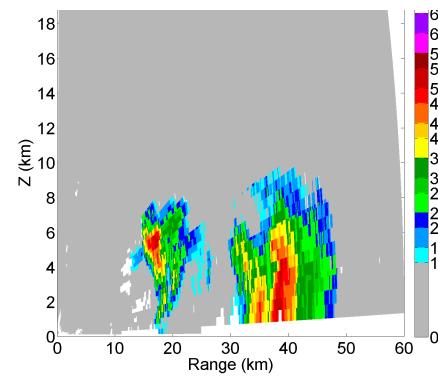
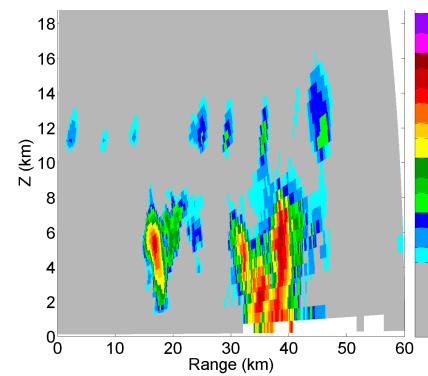
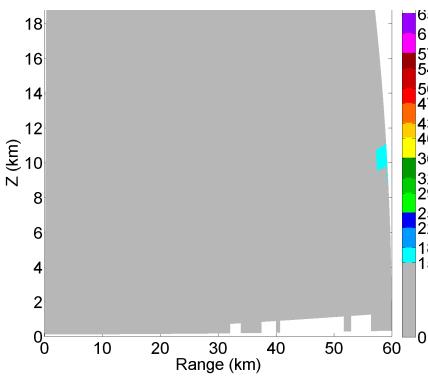
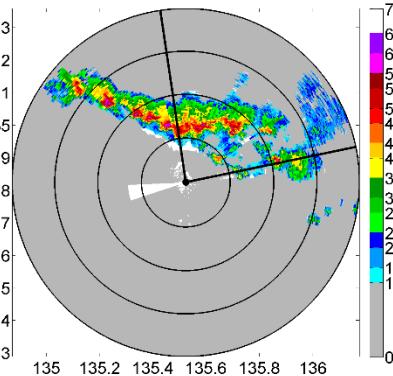
NO DA



WITH DA

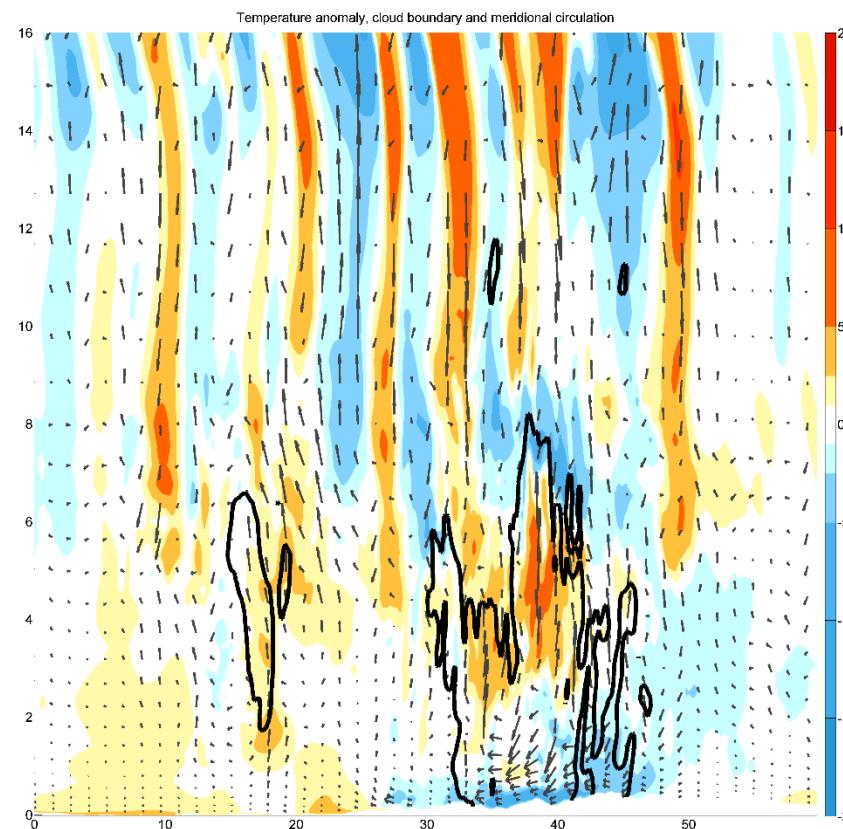


Observation



Thermodynamic structure

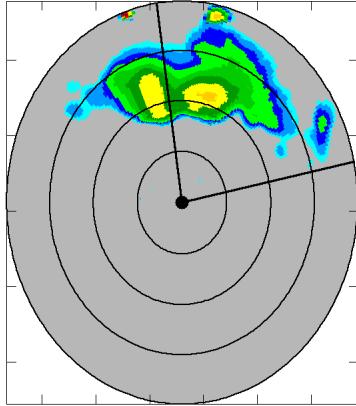
15:06:00 JST



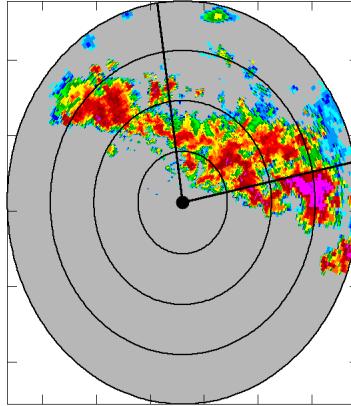
Forecasts

10-min. 15:06:00 JST → 15:16:00 JST

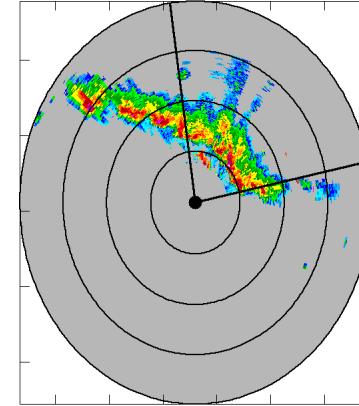
NO DA



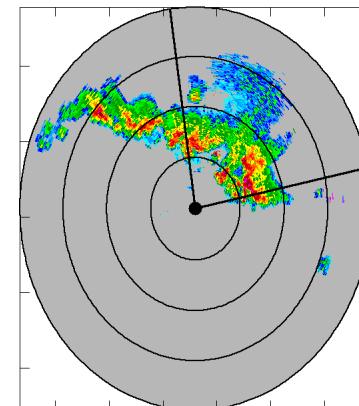
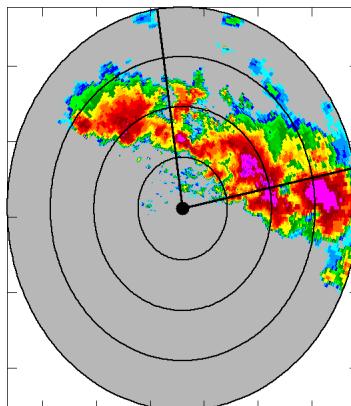
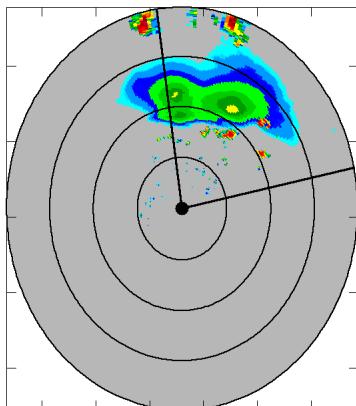
WITH DA



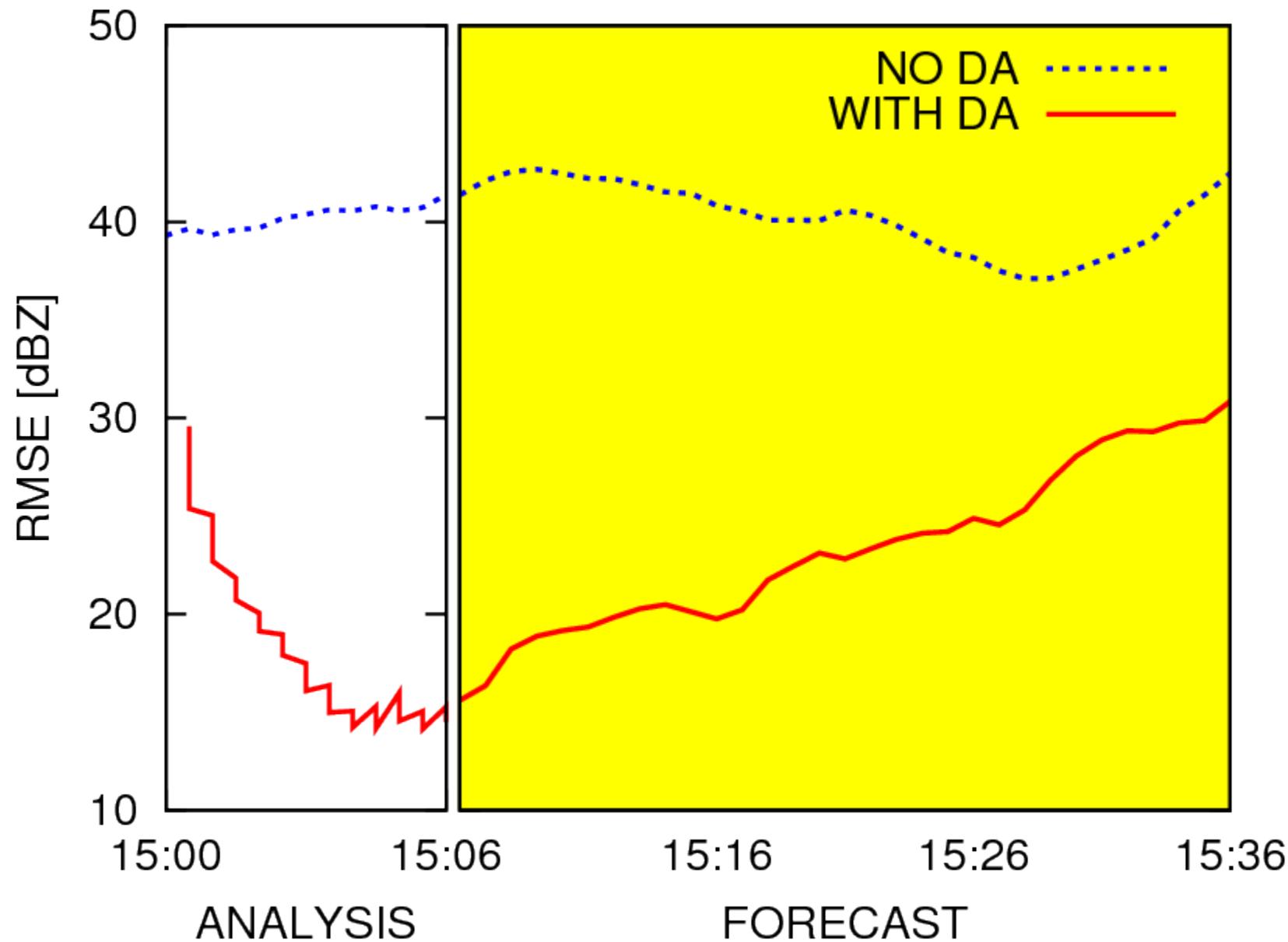
Observation



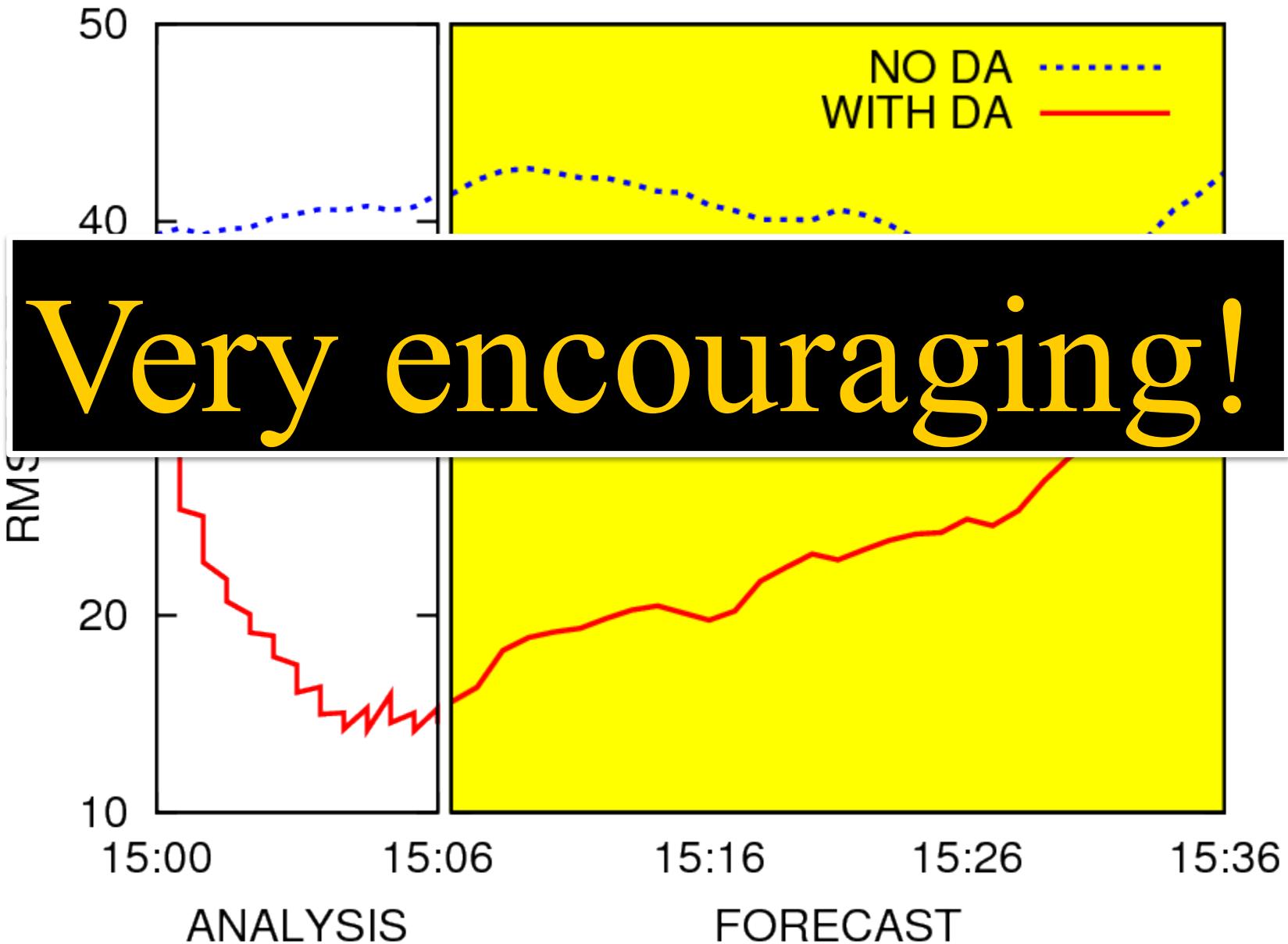
20-min. 15:06:00 JST → 15:26:00 JST



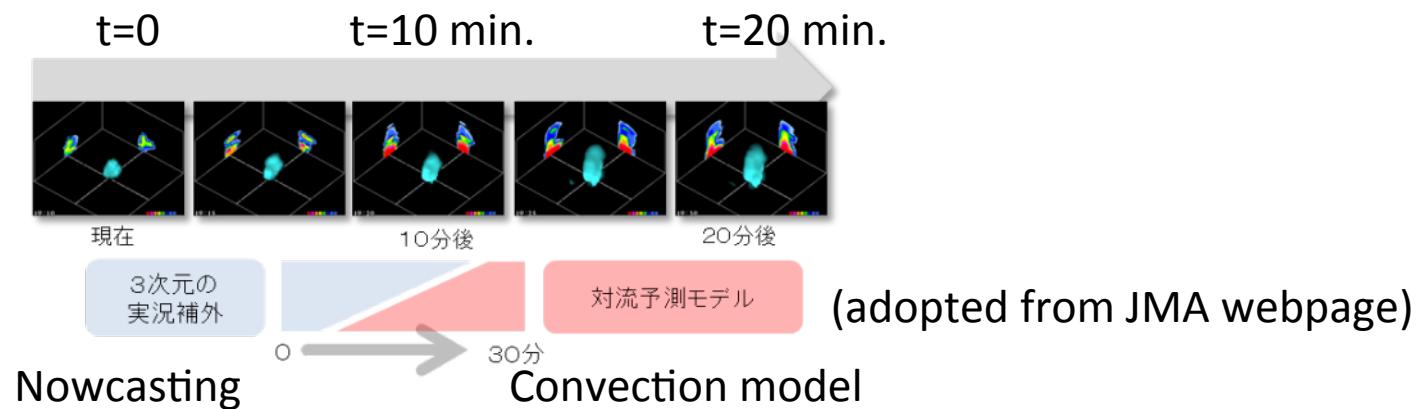
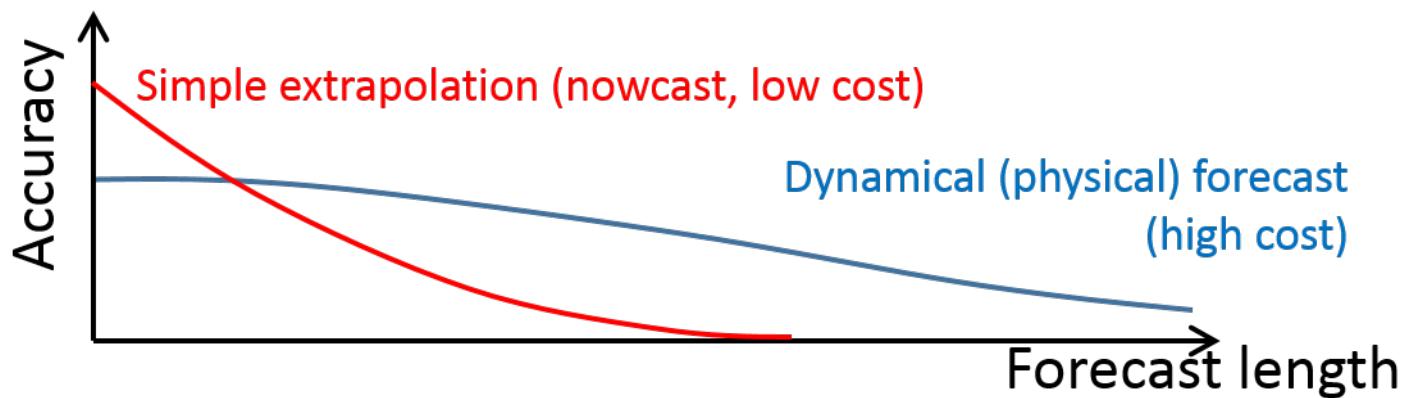
RMS Errors



RMS Errors



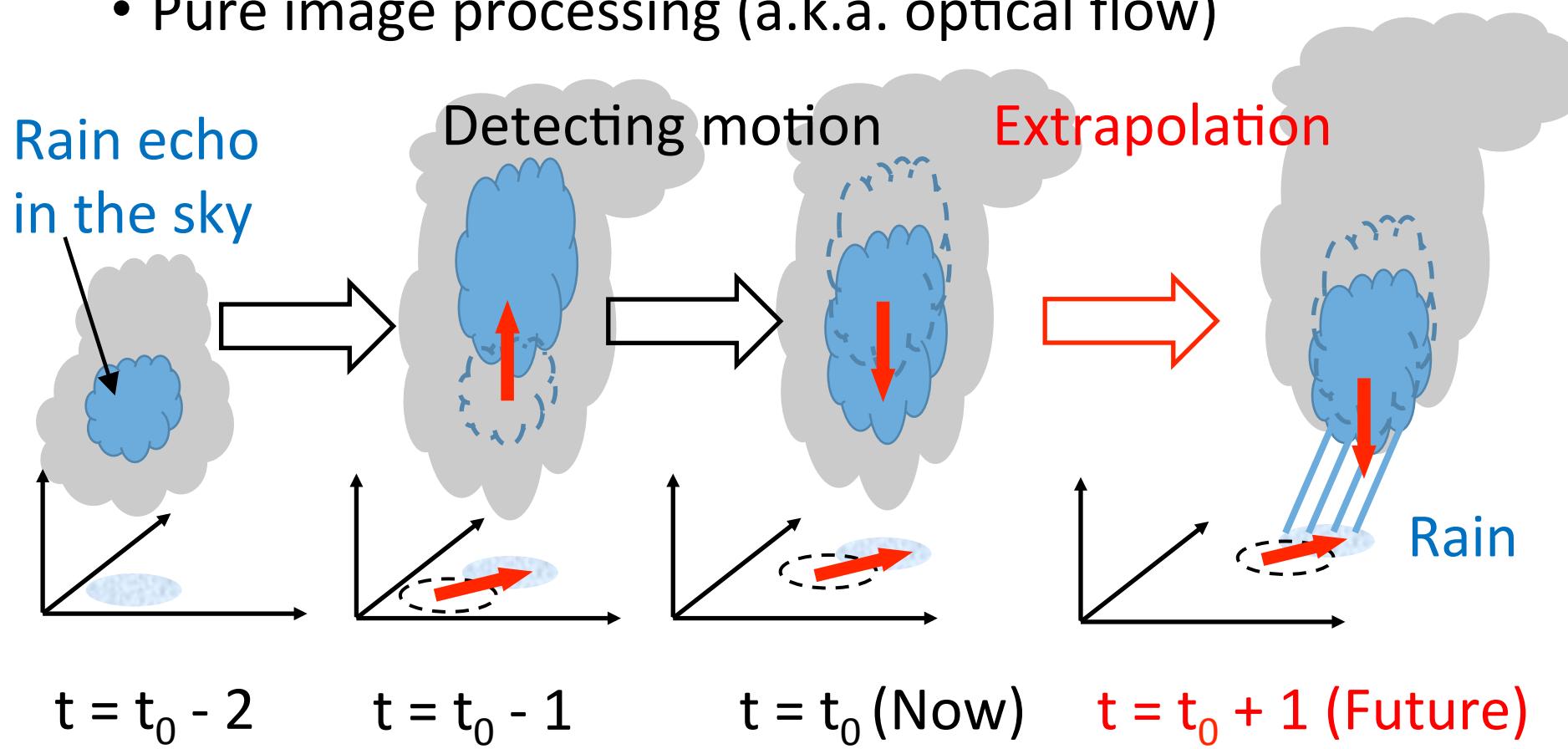
Toward seamless prediction



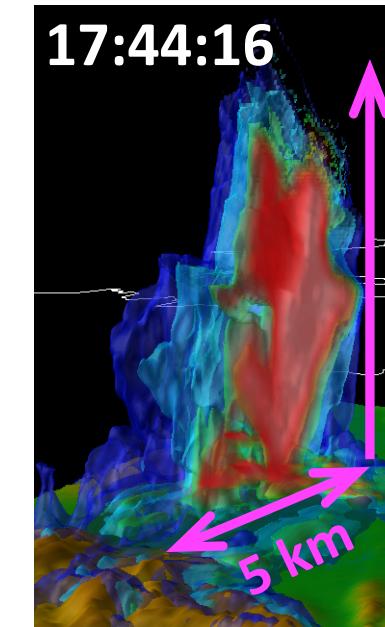
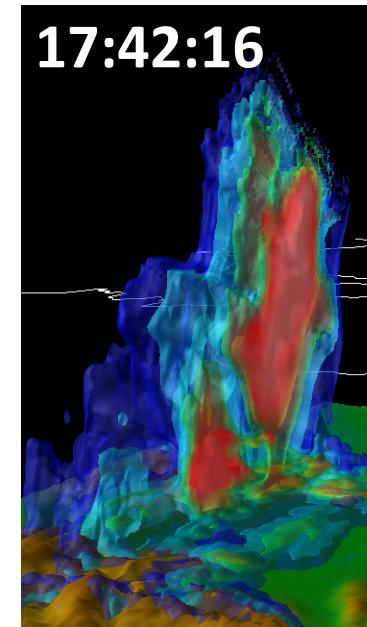
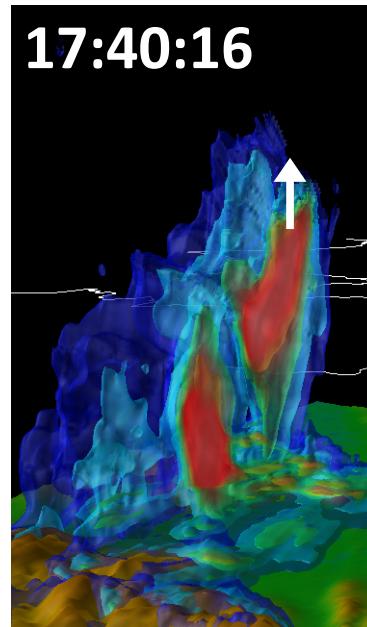
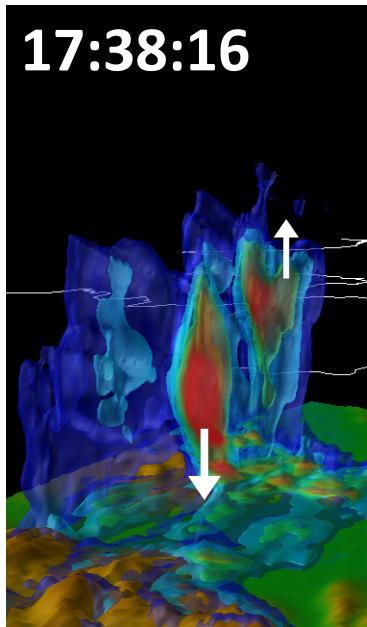
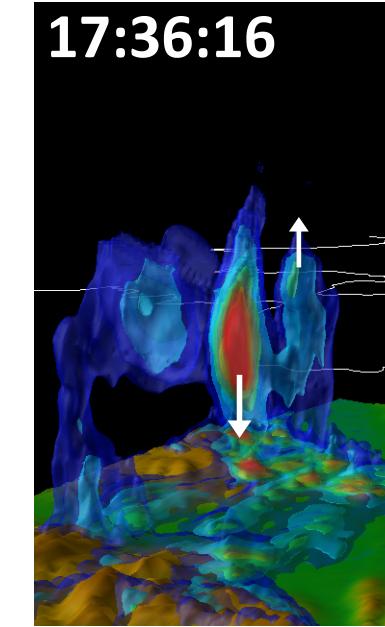
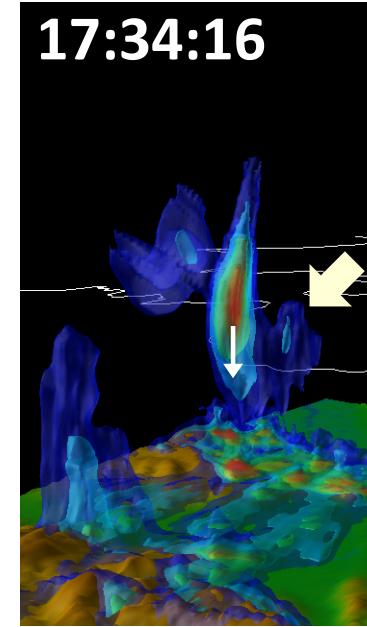
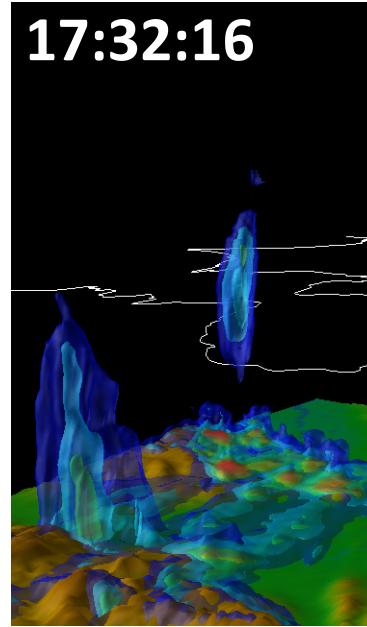
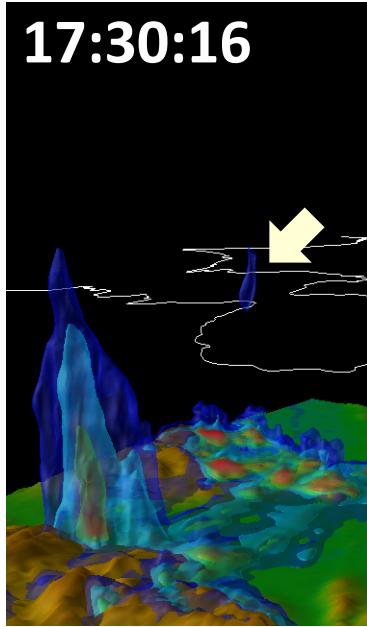
A nowcasting system taking advantage of the **dense/frequent** PAWR data is explored.

2D vs. 3D

- 3D motion extrapolation
- Pure image processing (a.k.a. optical flow)



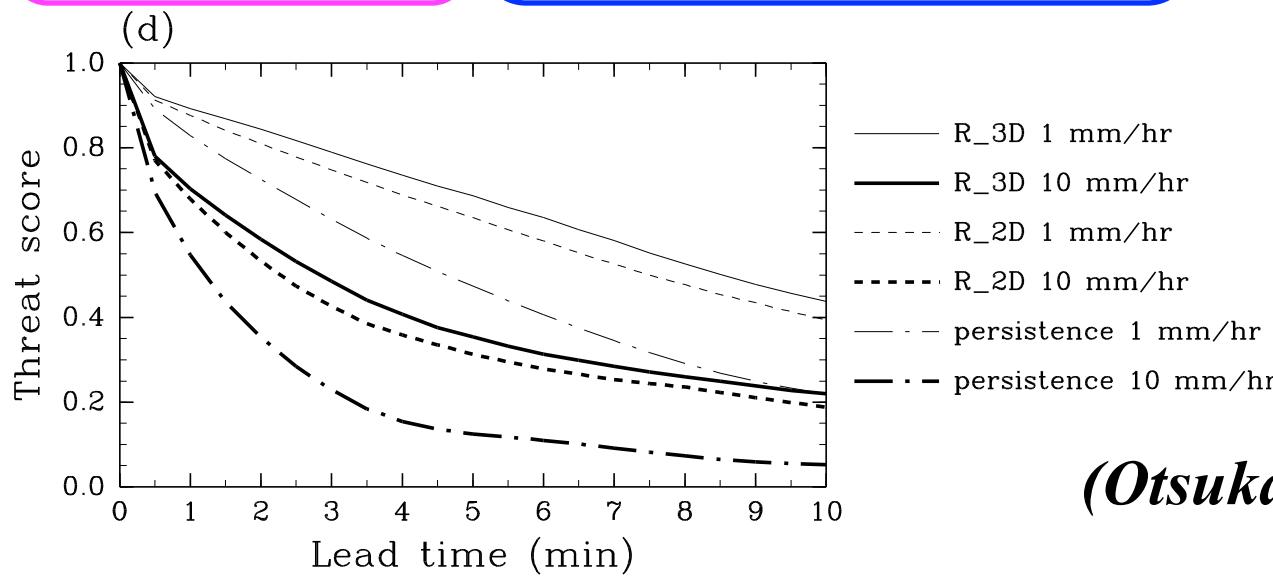
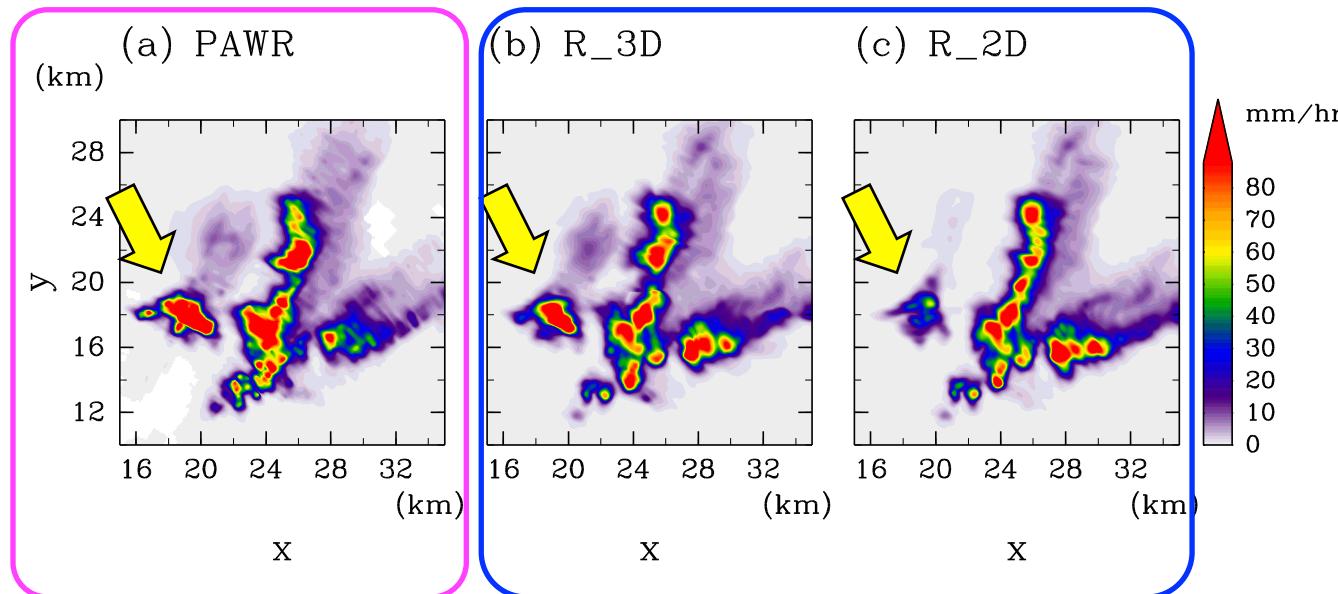
A real case: sky to the ground in 10 min.



With the **real** Phased Array Radar

observed

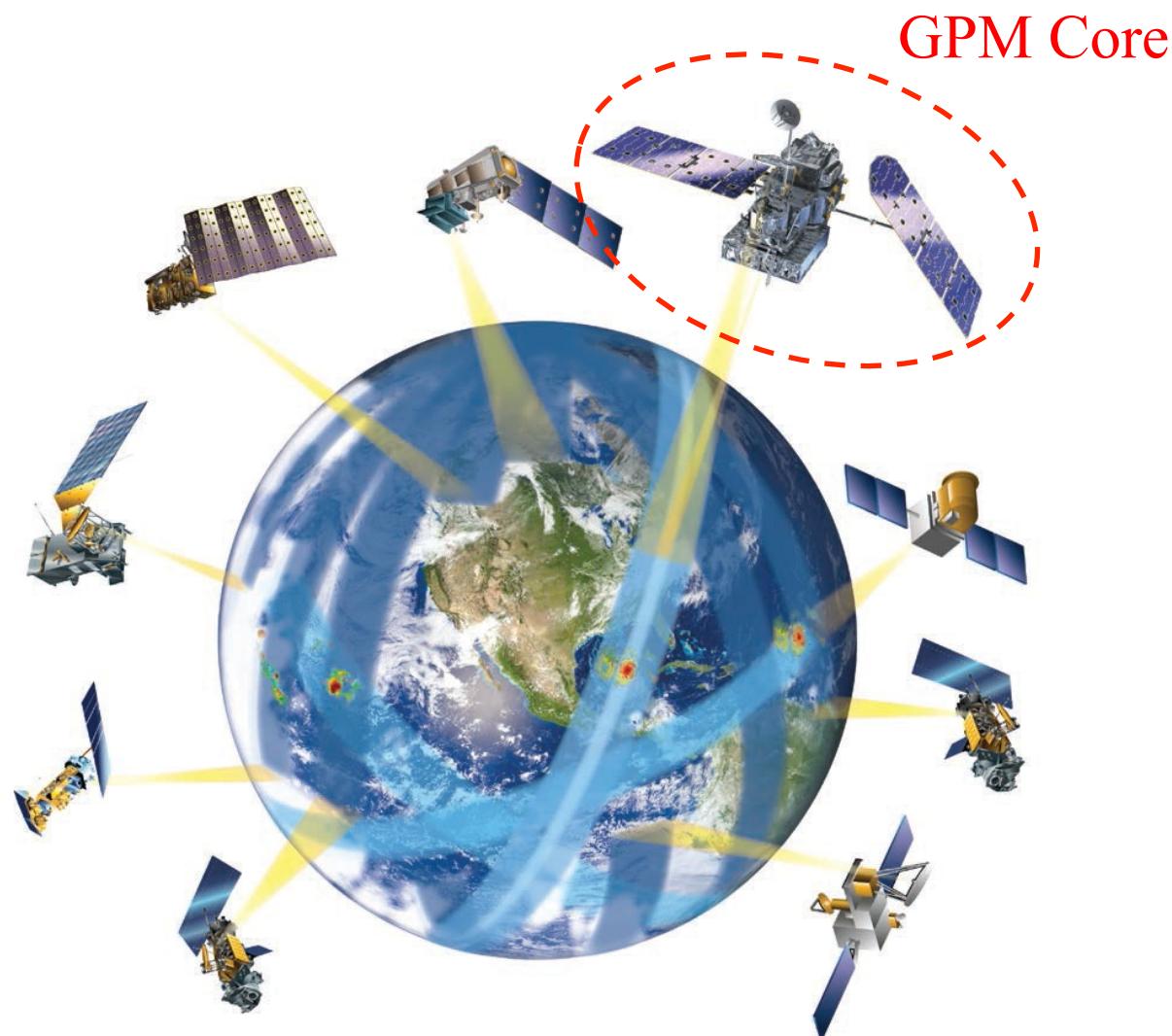
90-sec forecasts



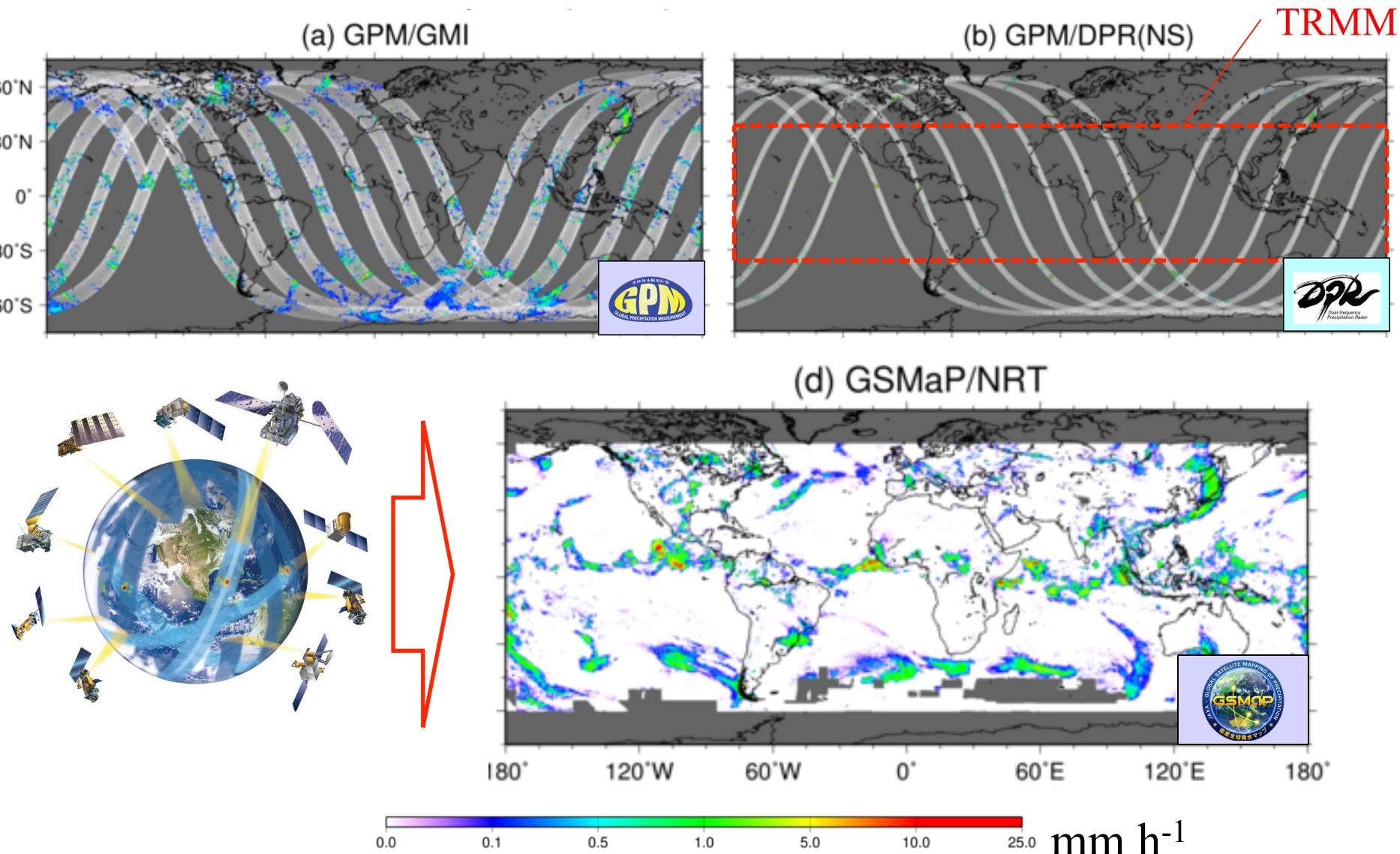
Underused satellite data potentially useful
PRECIPITATION

With Kotsuki, Terasaki, Lien, Otsuka, Tomita, Satoh, Kalnay

GPM: Global Precipitation Measurement



TRMM/GPM-derived precipitation



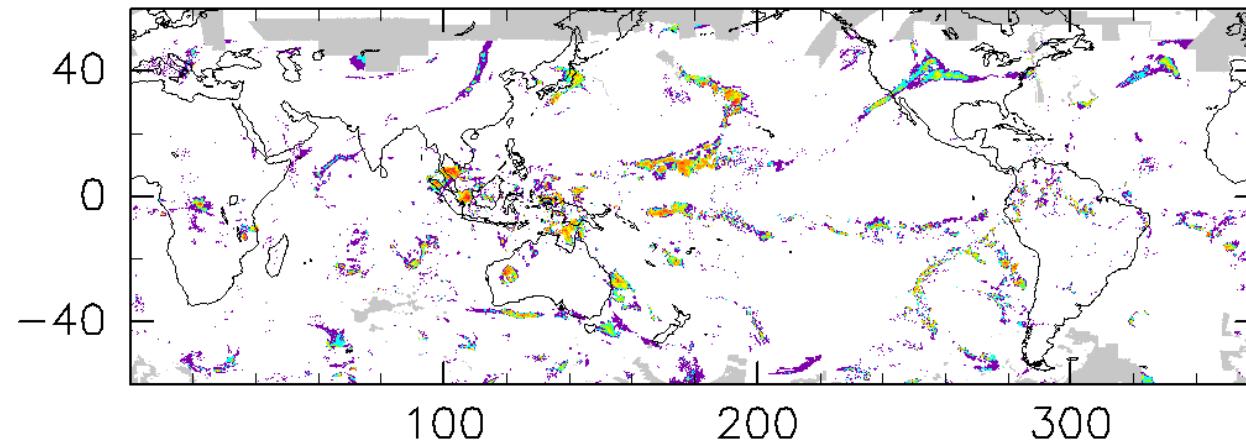
“Near Real Time” products (Ushio et al. 2009)

“Optical flow” (nowcasting)

2013/7/13 00Z

GSMaP

For “*Near Future*” products

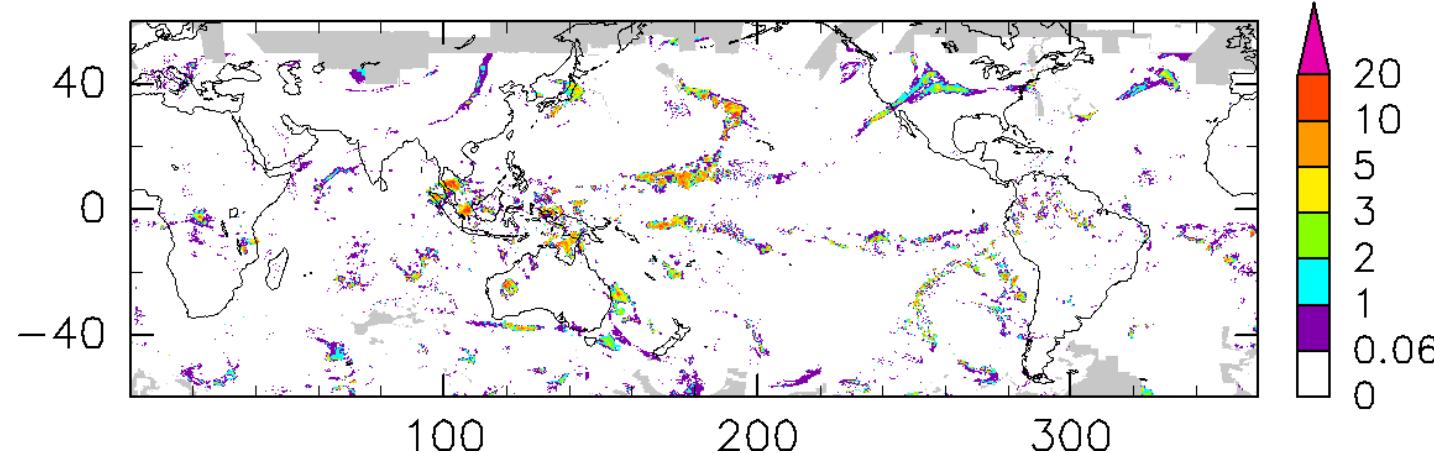


100

200

300

GSMaP



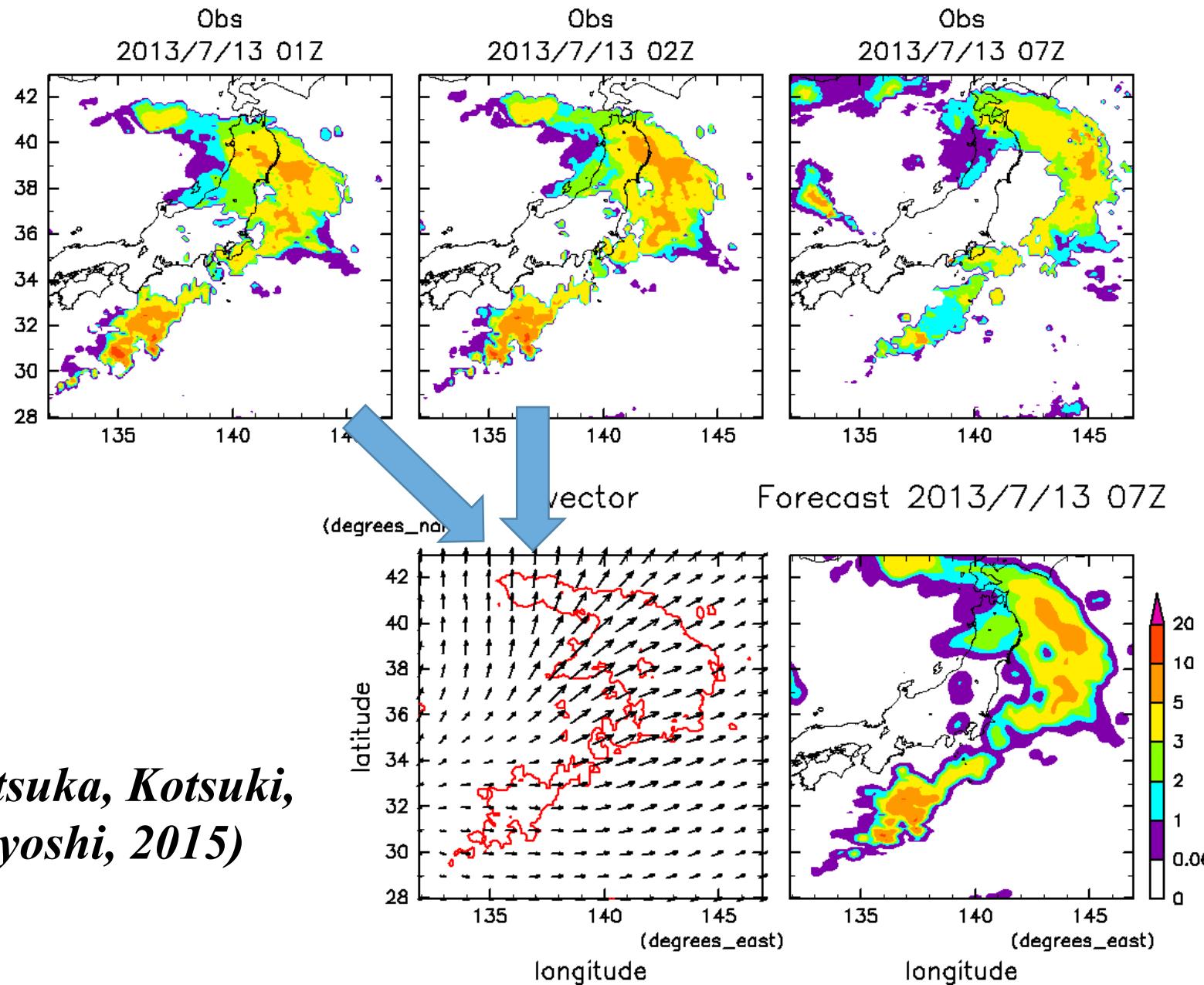
100

200

300

(Otsuka, Kotsuki, Miyoshi, 2015)

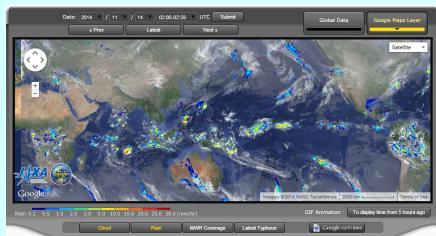
“Optical flow” (nowcasting)



*(Otsuka, Kotsuki,
Miyoshi, 2015)*

GSMaP assimilated into NICAM

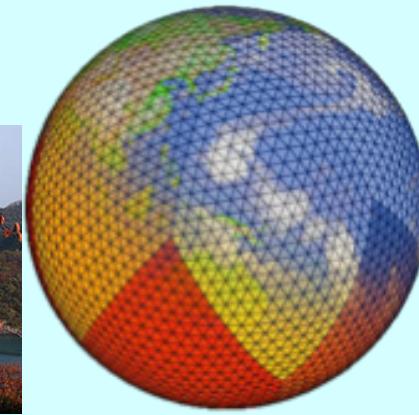
GSMaP_NRT



LETKF



NICAM

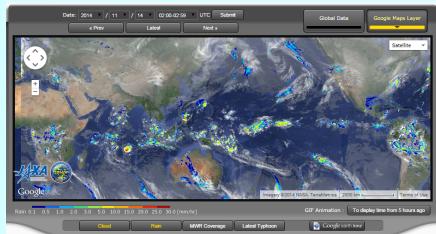


Local Ensemble Transform Kalman Filter
(Hunt et al. 2007)

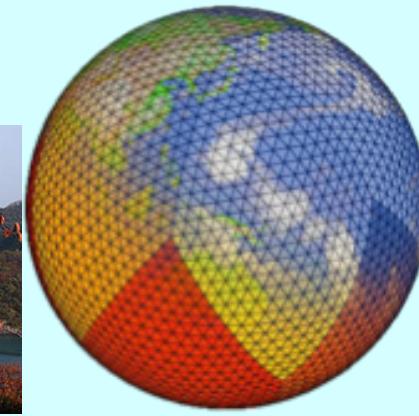
Toward seamless
NRT → Now → Near Future → Forecast

GSMaP assimilated into NICAM

GSMaP_NRT



NICAM



LETKF



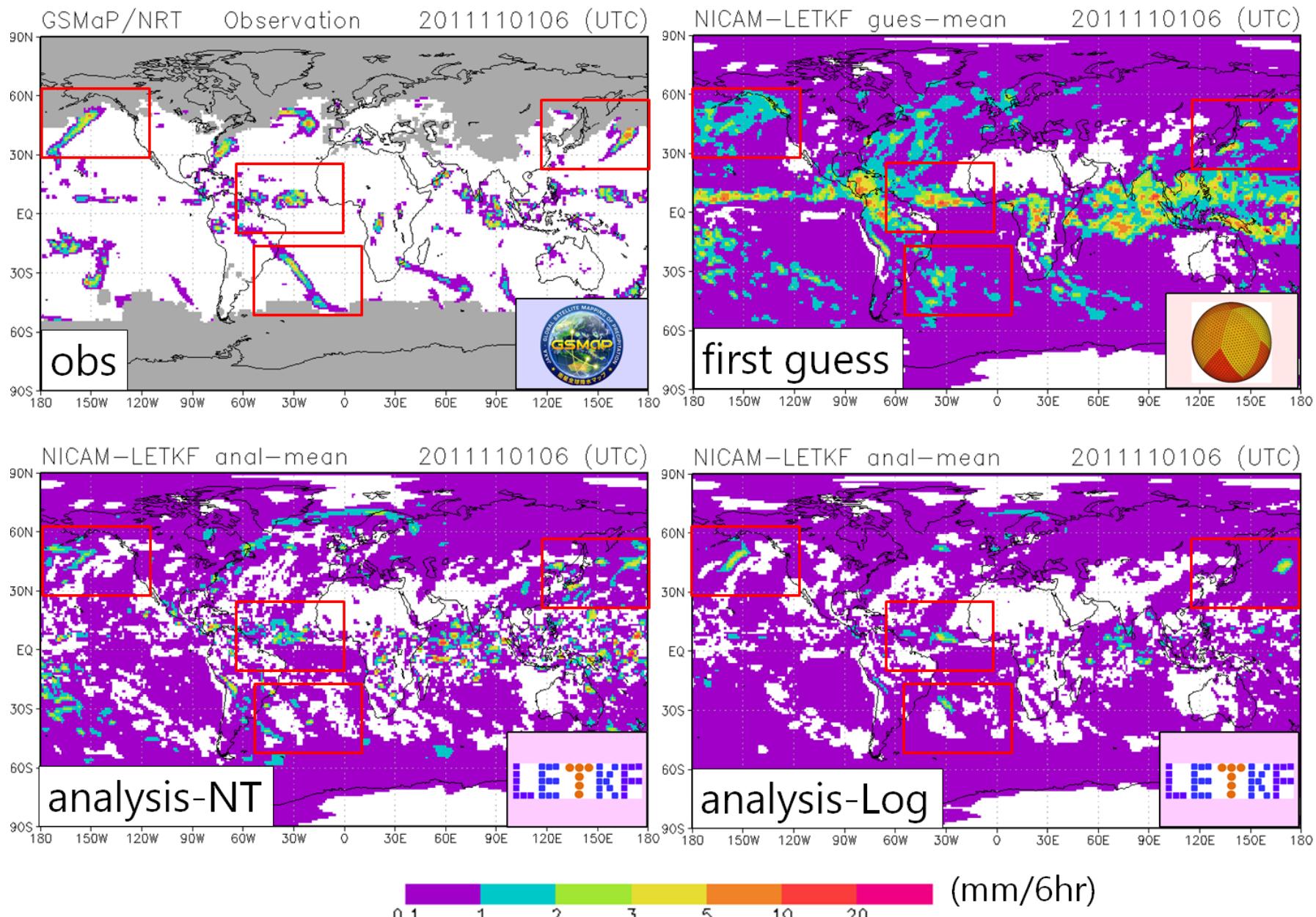
Local Ensemble Transform Kalman Filter

(Hunt et al. 2007)

**Toward seamless
NRT → Now → Near Future → Forecast**

**(Re-)Analysis = Satellite Level-4 products?
(an idea suggested by M. Satoh)**

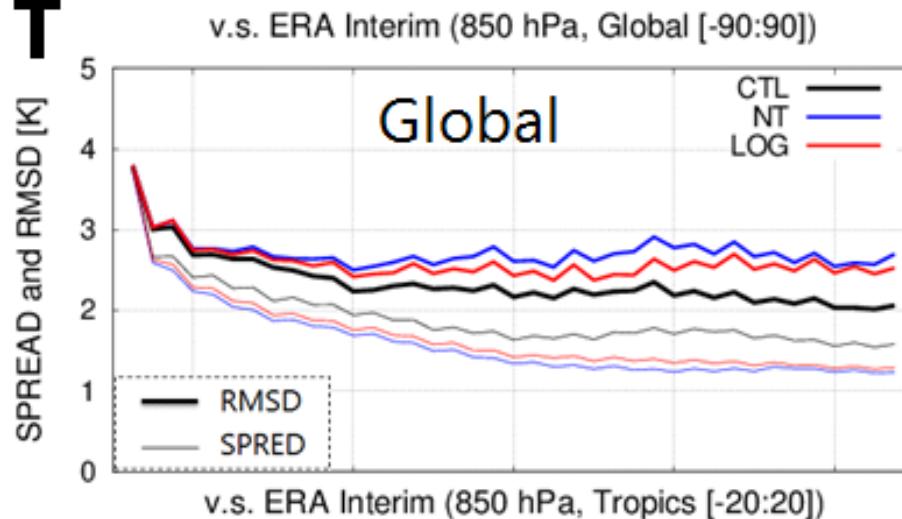
Unsuccessful attempt of assimilating raw or log-transformed precipitation *(Kotsuki et al., 2015)*



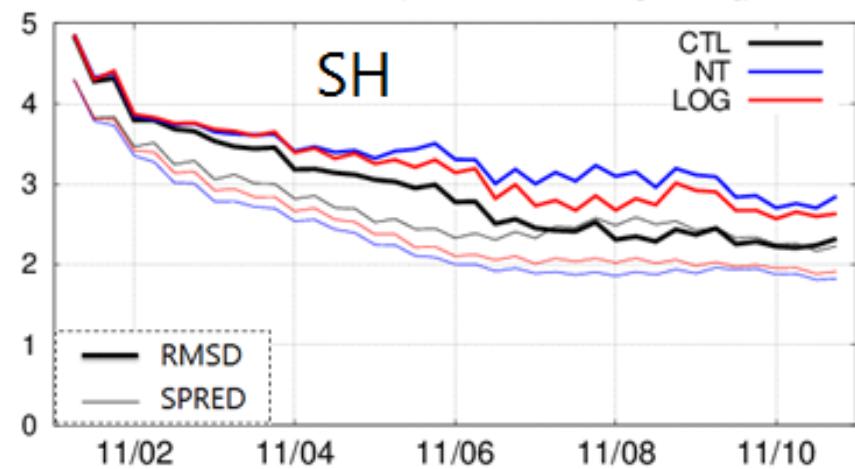
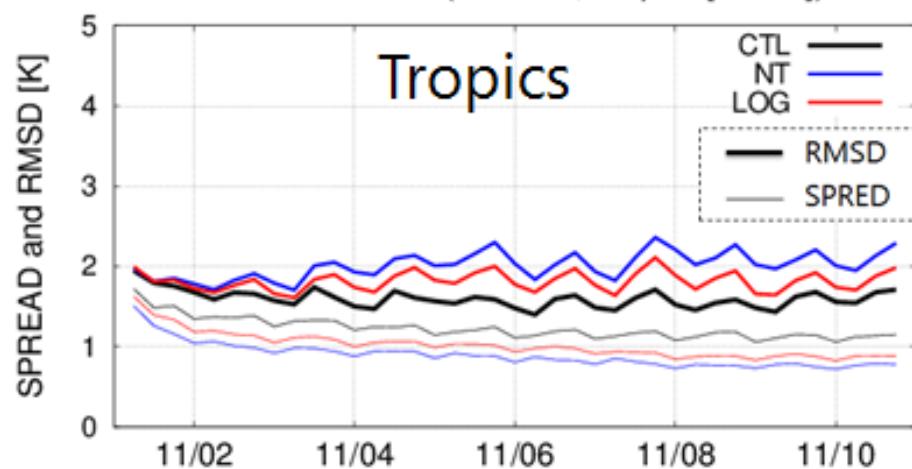
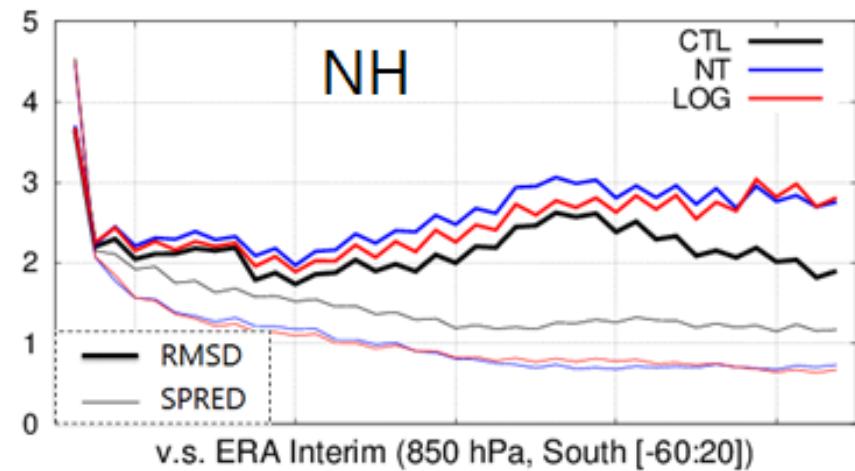
RMSD vs. ERA-int degraded due to precip. DA

(Kotsuki et al., 2015)

T



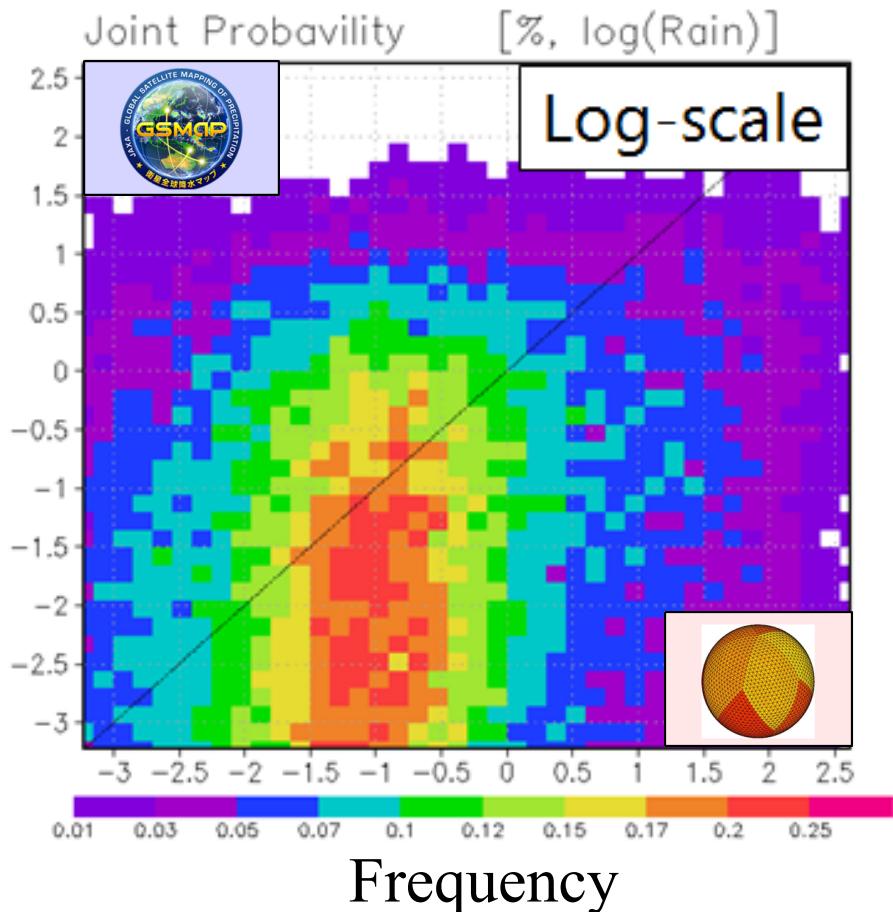
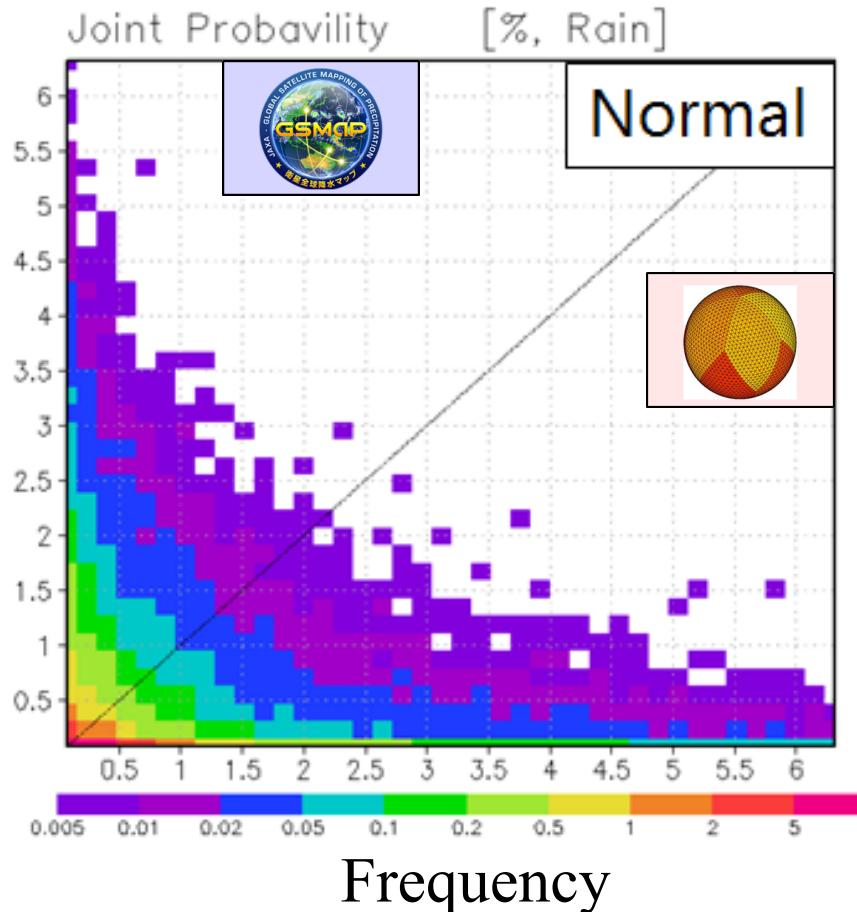
v.s. ERA Interim (850 hPa, North [20:60])



no assimilation > Log-transformation > no-transformation

Joint histograms [%]

(Kotsuki et al., 2015)

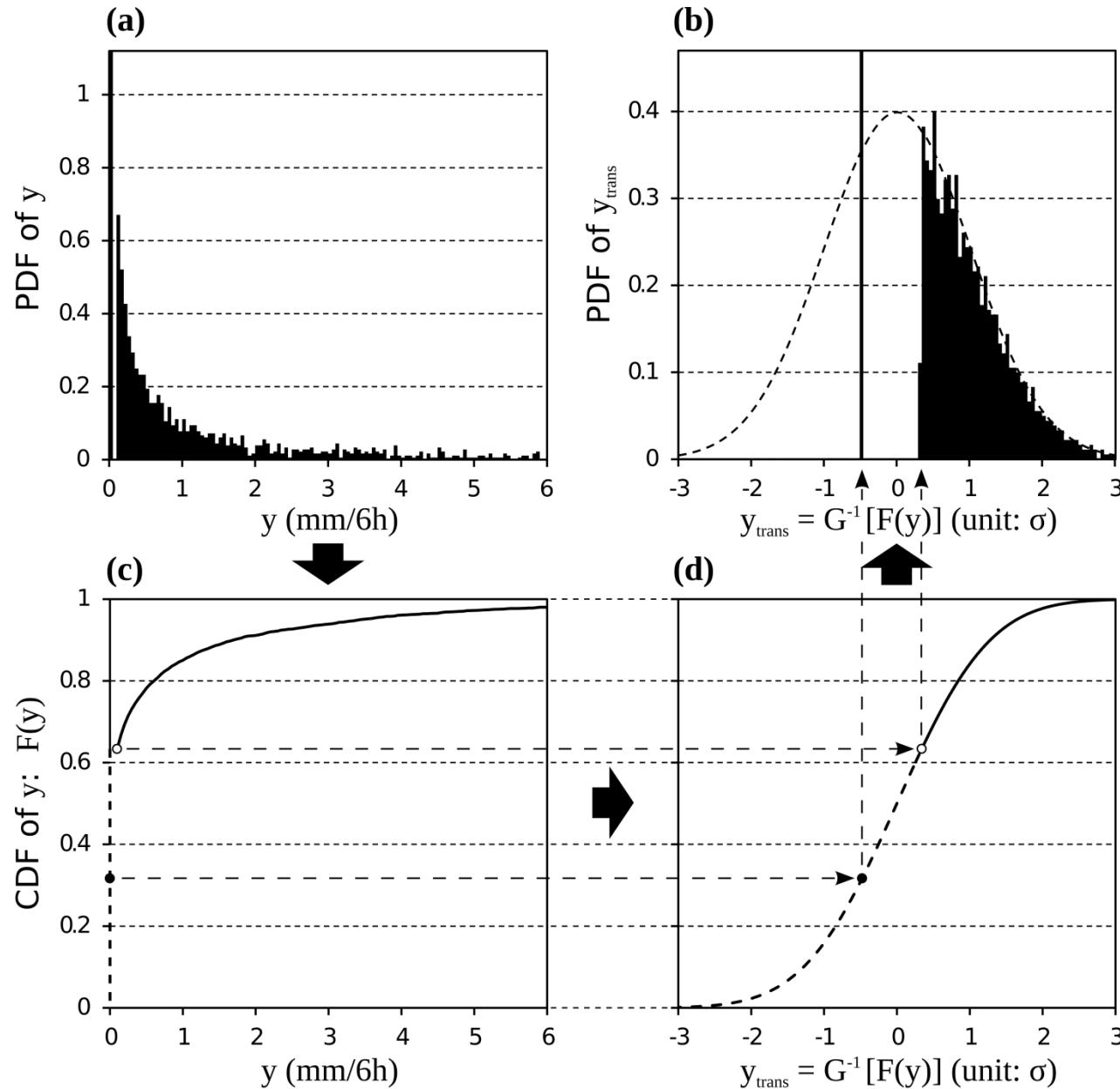


→ Gaussian transformation?

RECENT SUCCESS

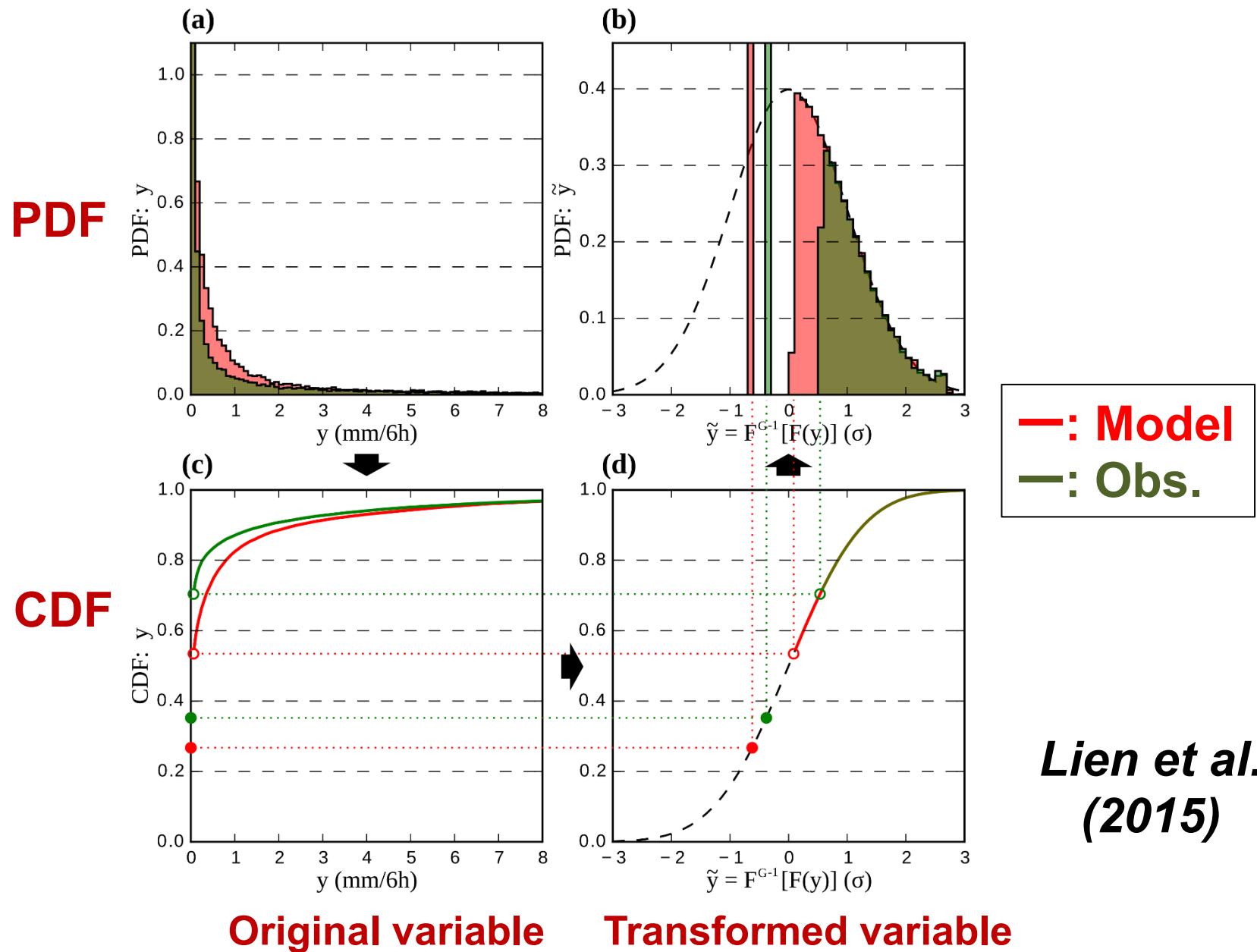
Lien, Miyoshi, Kalnay (2013, *Tellus*)
Lien et al. (2015, in revision)
Hotta et al. (2015, in prep.)

Example of precipitation distribution in DJF near Maryland

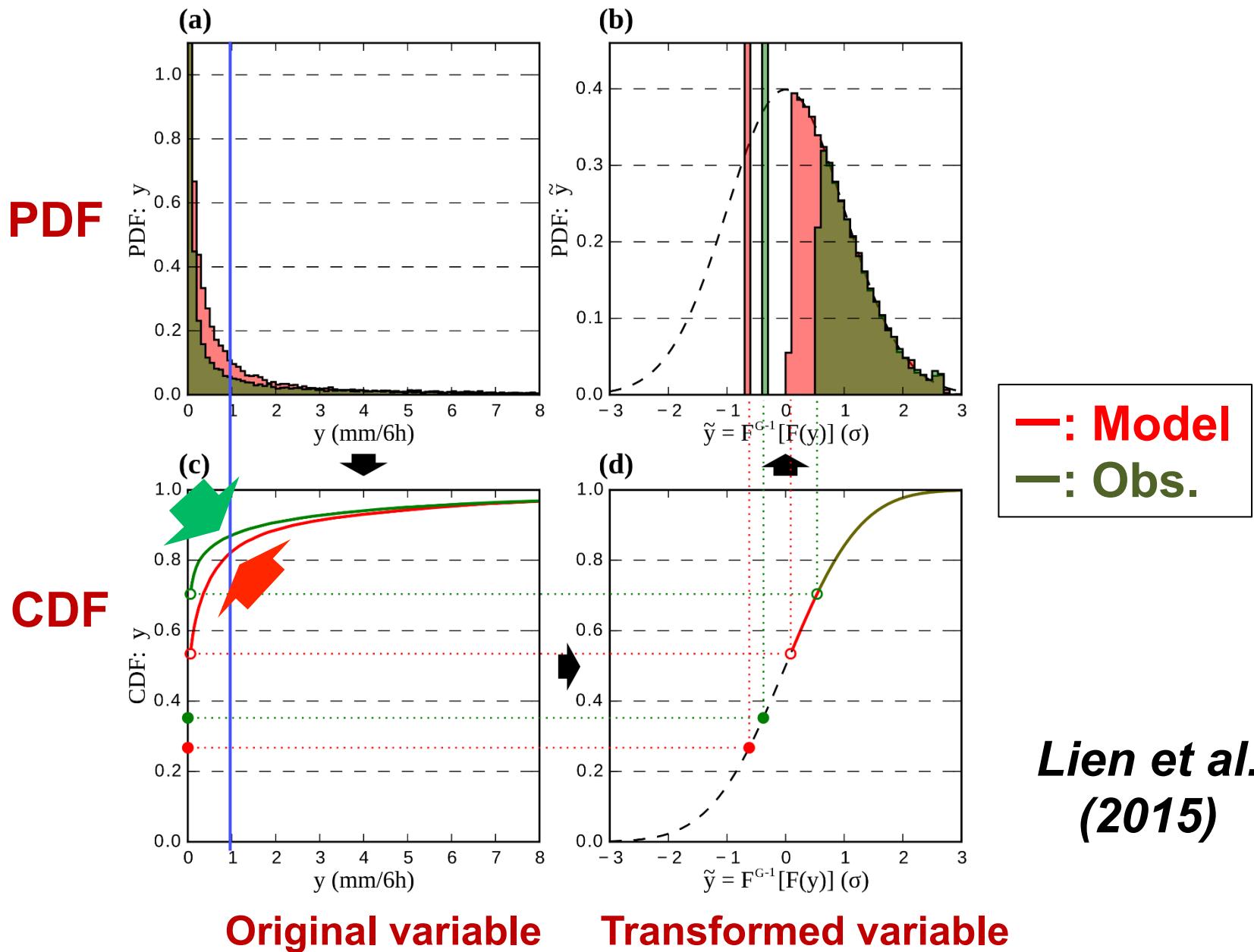


*Lien et al.
(2013)*

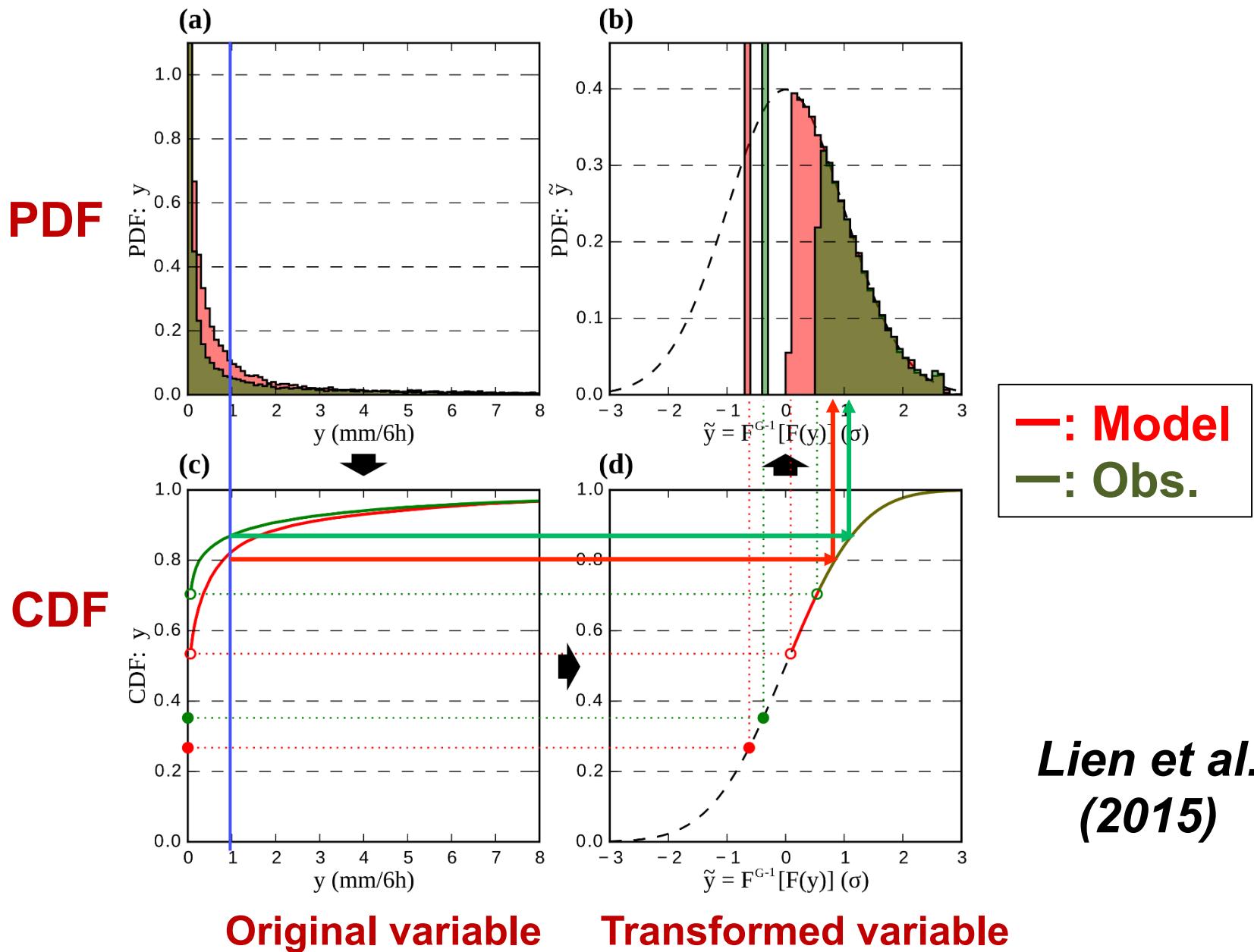
Gaussian Transformation for the real case



Gaussian Transformation for the real case



Gaussian Transformation for the real case

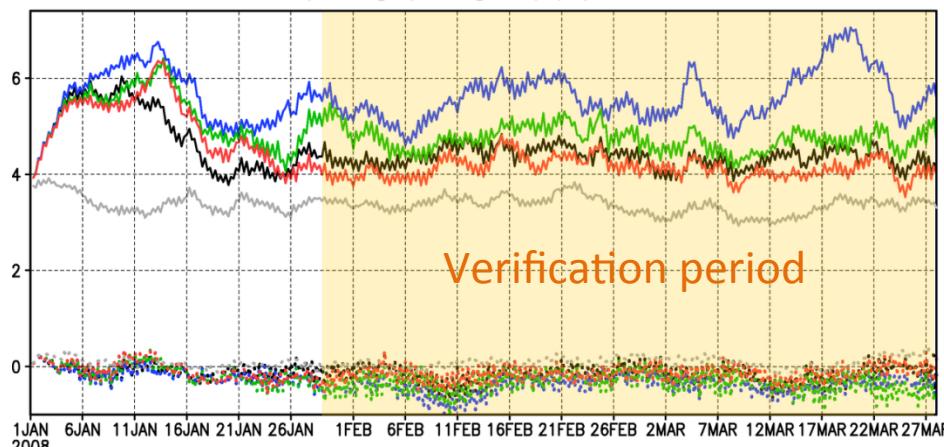


Lien and Kalnay succeeded in assimilating TMPA.

3-month time series: Analysis U (m/s) at 500 hPa, error relative to ERA-int.

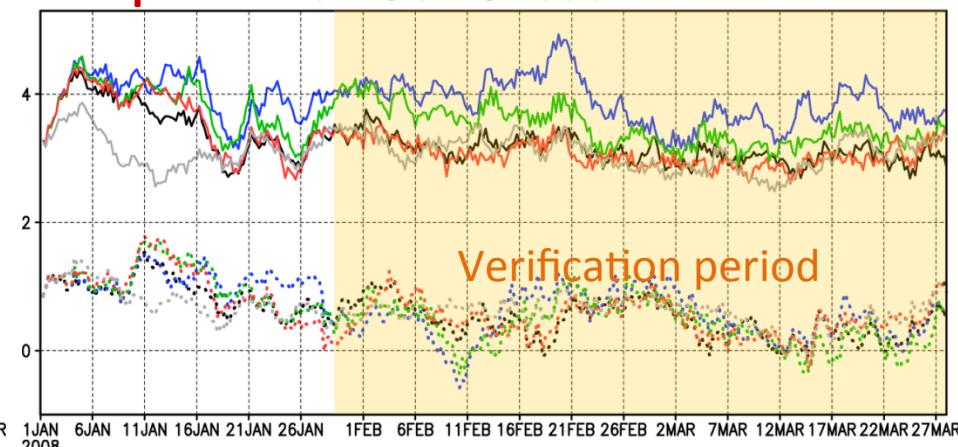
Global

RMSE/Bias [GL/t=0h]: U (m/s) at 500hPa



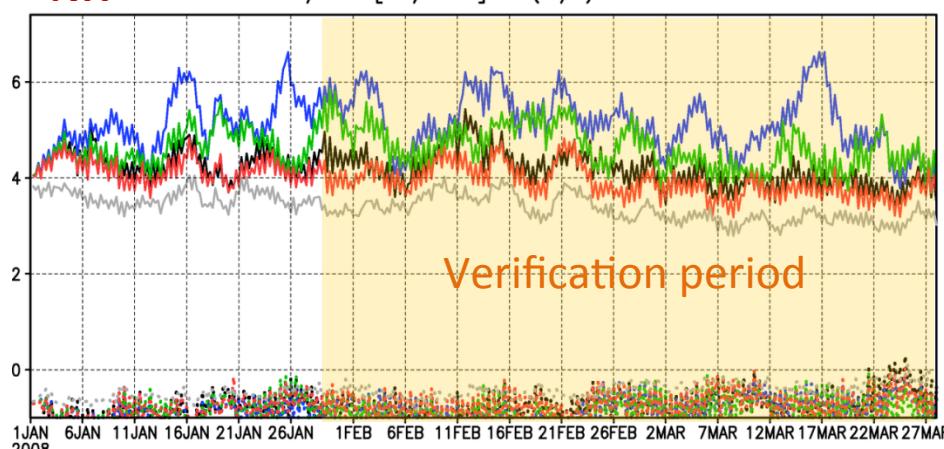
Tropics

RMSE/Bias [TR/t=0h]: U (m/s) at 500hPa



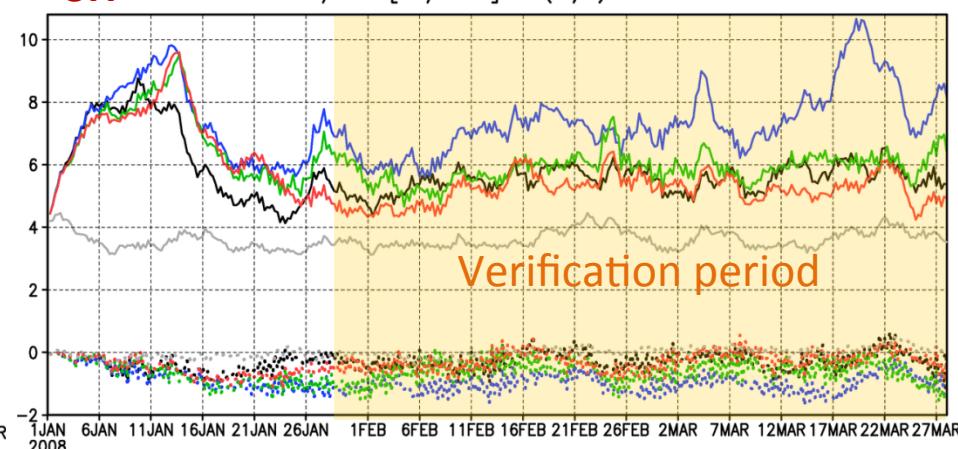
NH

RMSE/Bias [NH/t=0h]: U (m/s) at 500hPa



SH

RMSE/Bias [SH/t=0h]: U (m/s) at 500hPa



Raobs

Raobs_TMPA

Raobs_TMPA_Log

Raobs_TMPA_GT

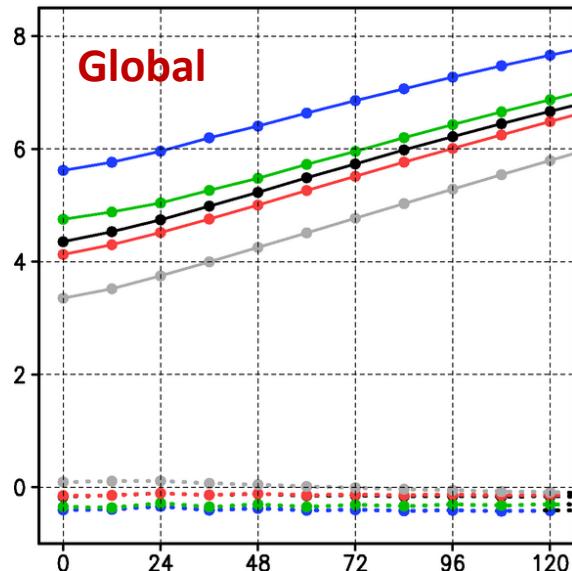
Conv

Solid lines: RMS error

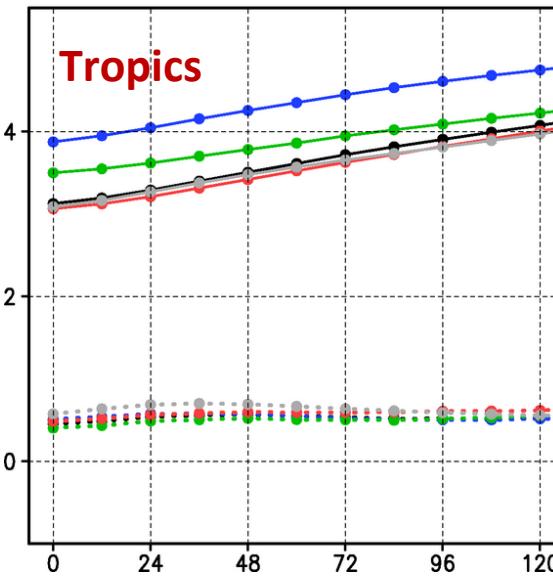
Dashed lines: Bias

5-day forecasts were improved.

RMSE/Bias [GL]: U (m/s) at 500hPa



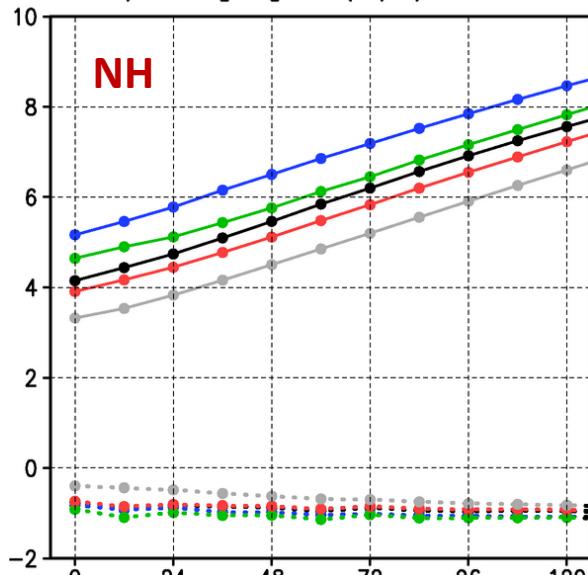
RMSE/Bias [TR]: U (m/s) at 500hPa



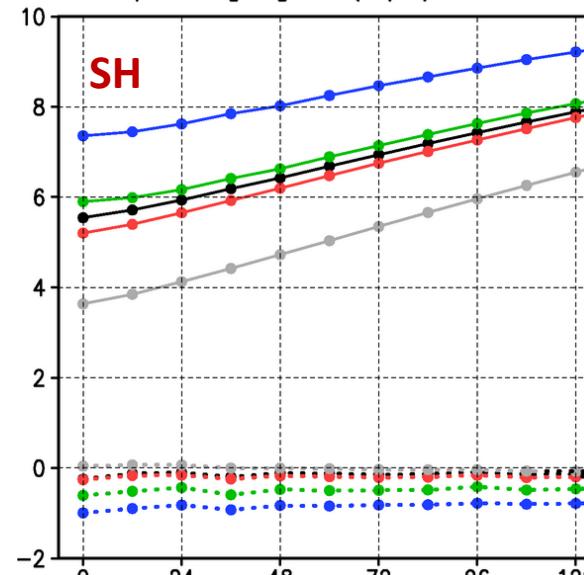
2-month average
U at 500 hPa

Solid lines: RMS error
Dashed lines: Bias

RMSE/Bias [NH]: U (m/s) at 500hPa



RMSE/Bias [SH]: U (m/s) at 500hPa

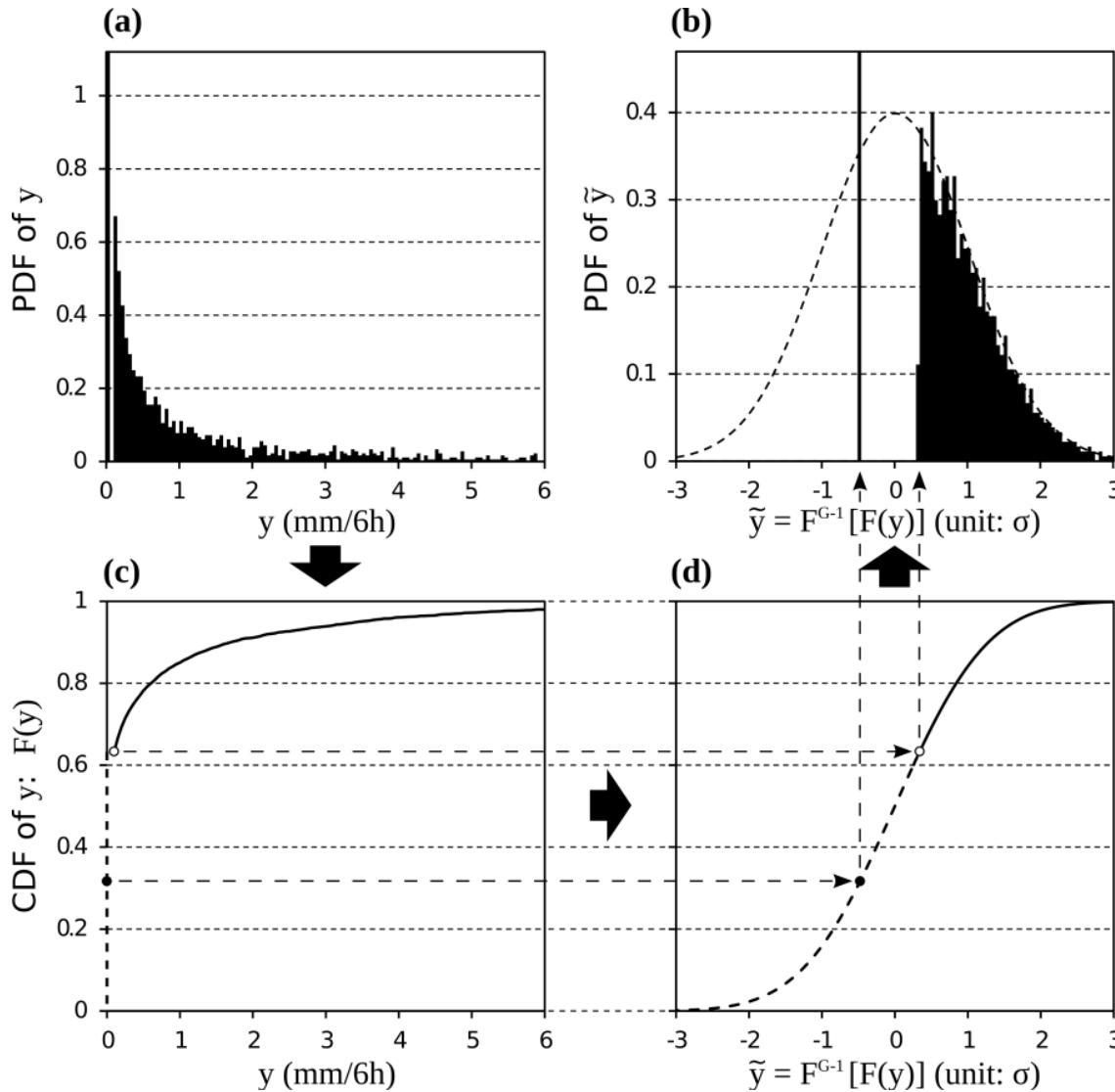


- Raobs
- Raobs_TMPA
- Raobs_TMPA_Log
- Raobs_TMPA_GT
- Conv

APPLYING TO NICAM-LETKF

With Kotsuki, Terasaki, Lien

Gaussian transformation



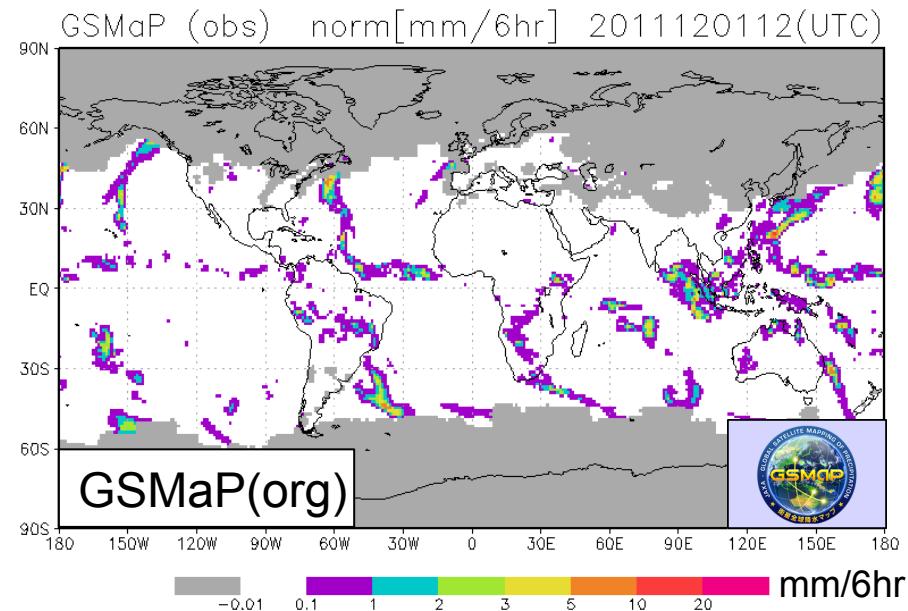
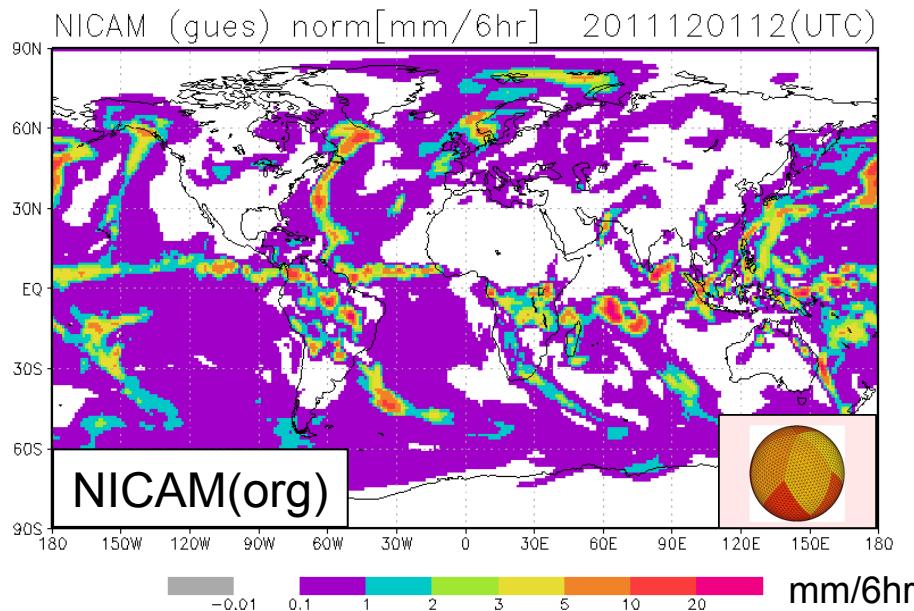
Precipitation
(mm/hr)

Gaussian Transformation

Gaussian Var.
(sigma)

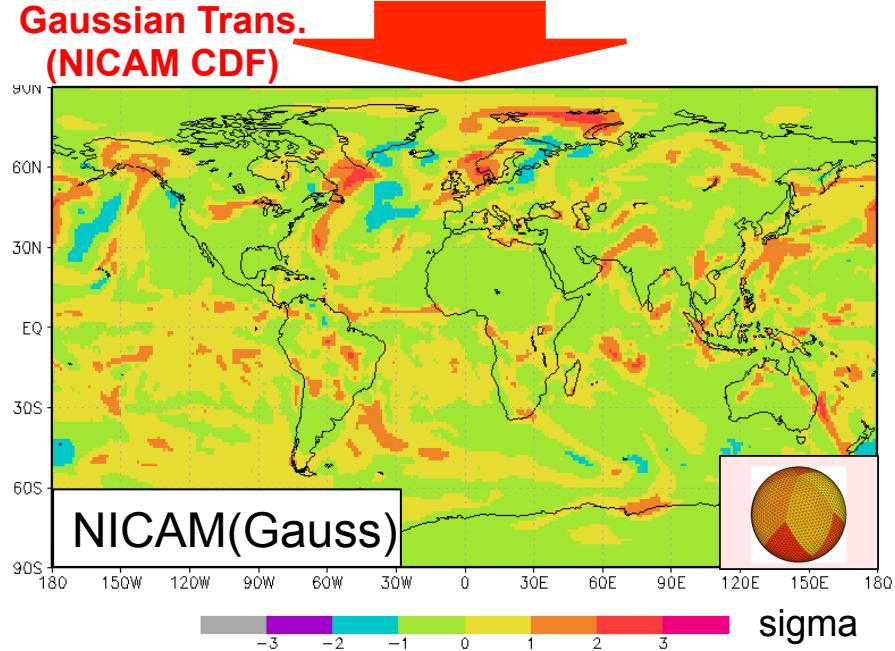
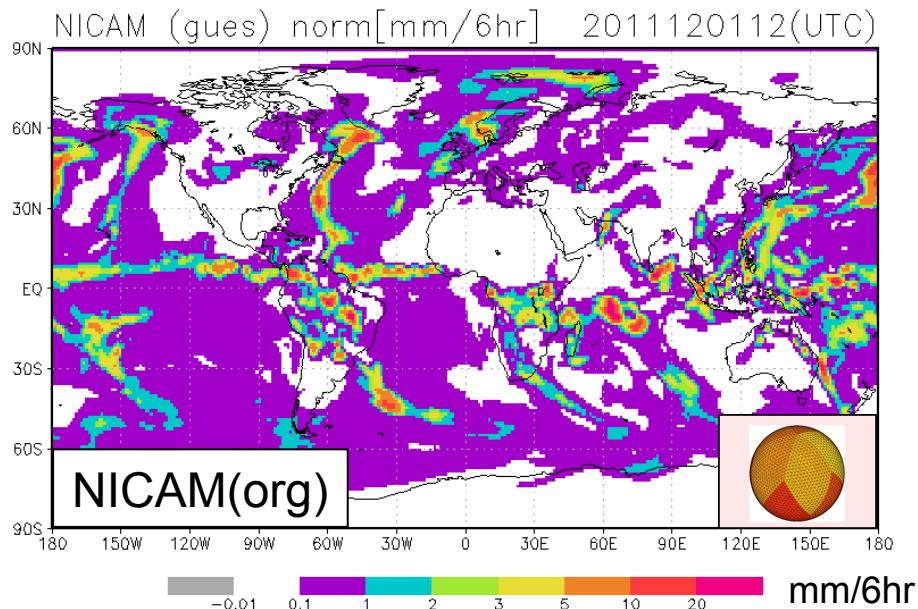
Gaussian transformation

(Kotsuki et al., 2015)

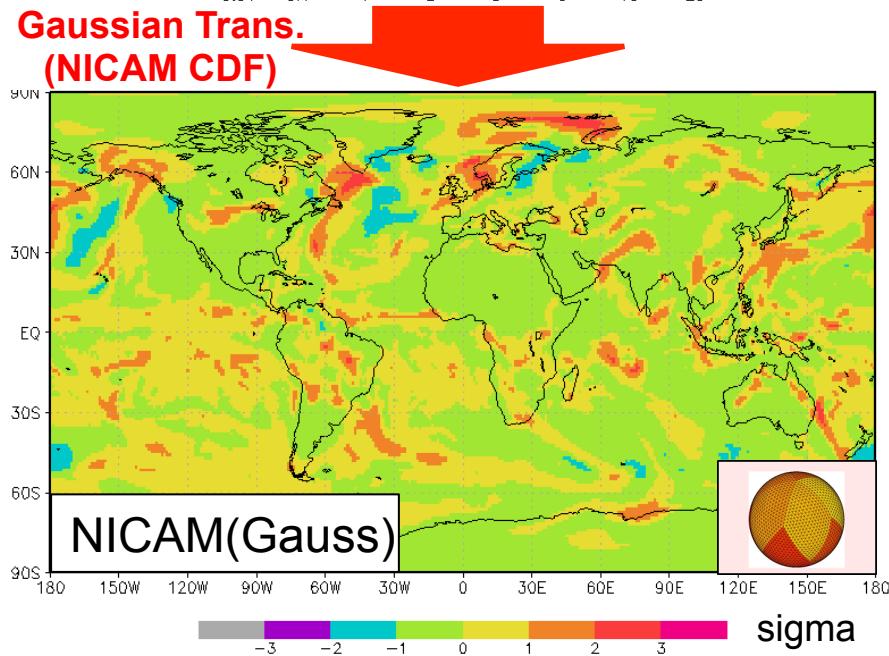
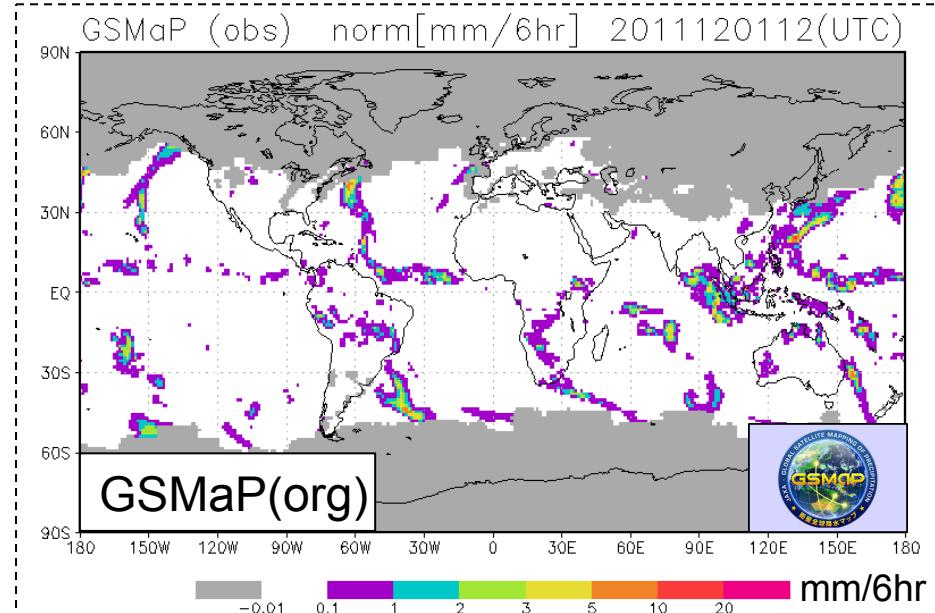
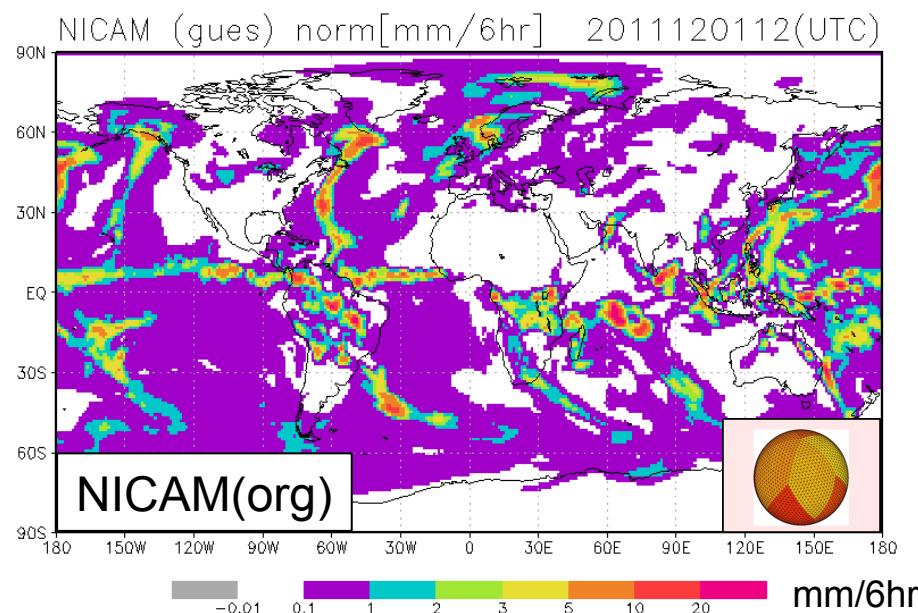


Gaussian transformation

(Kotsuki et al., 2015)

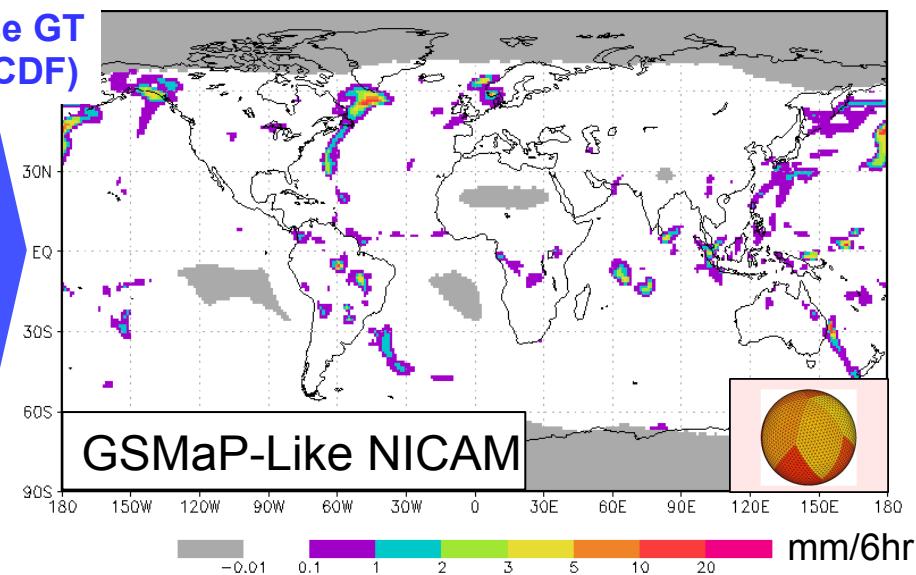
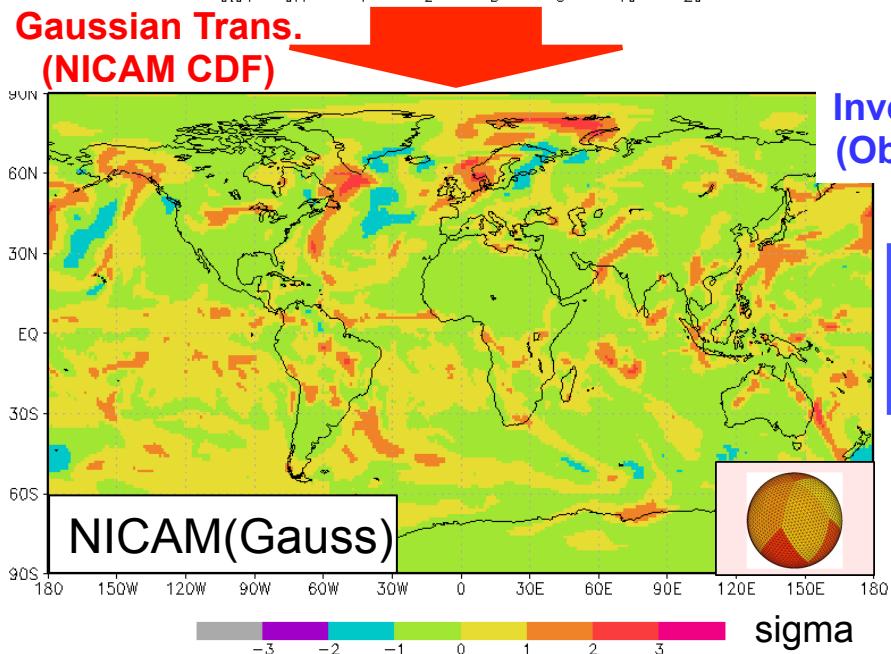
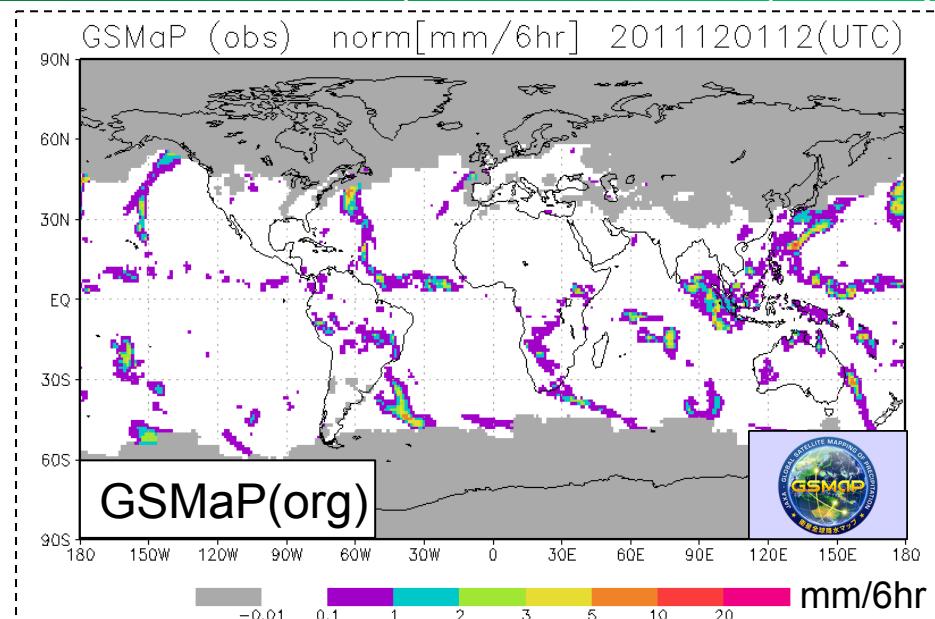
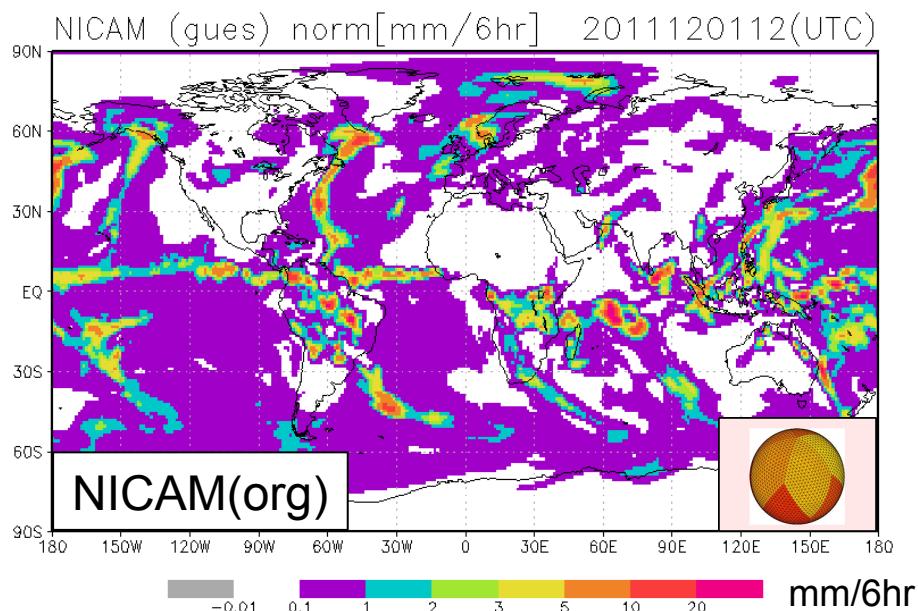


GSMaP-like NICAM using GT *(Kotsuki et al., 2015)*

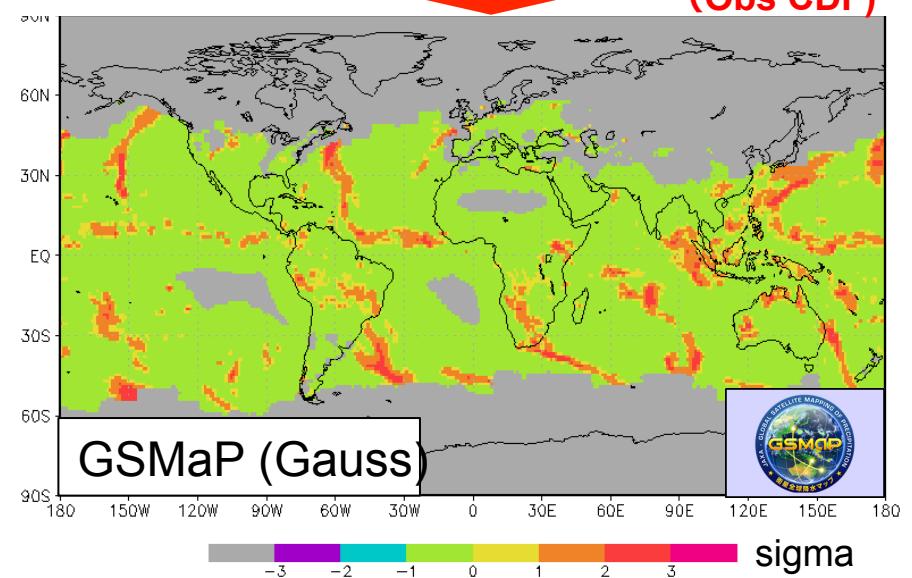
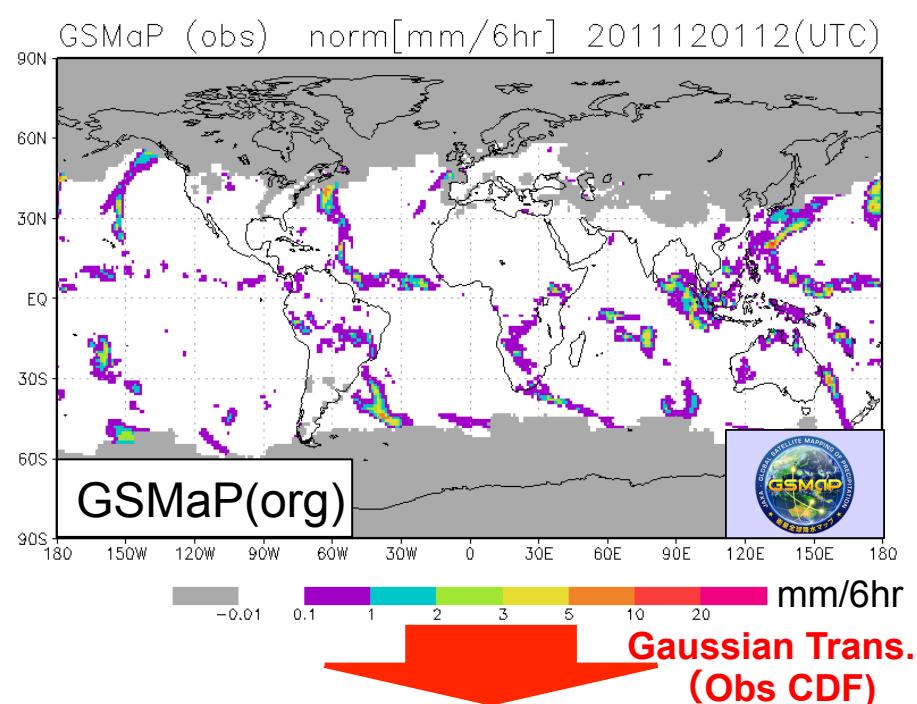
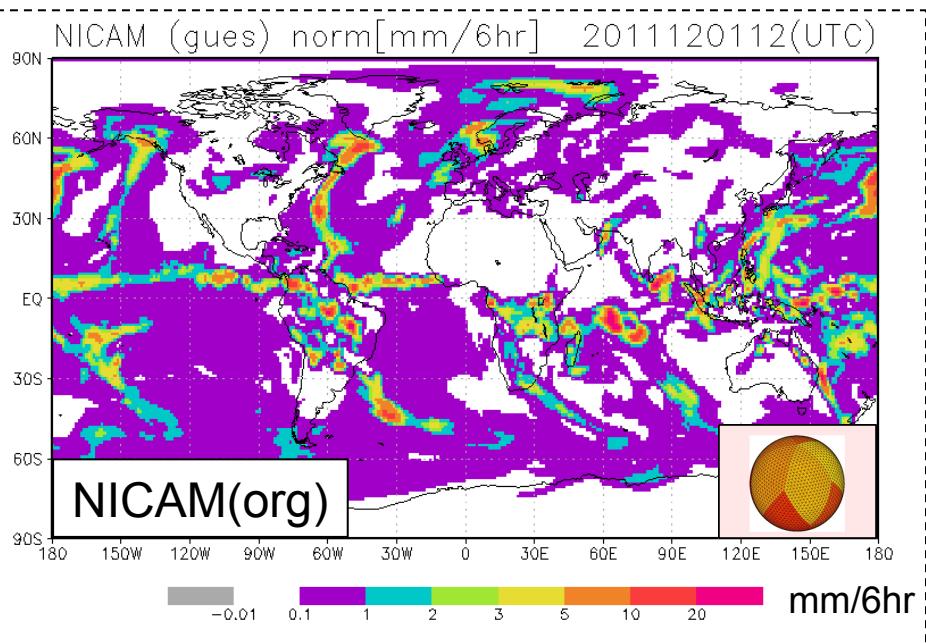


GSMaP-like NICAM using GT

(Kotsuki et al., 2015)

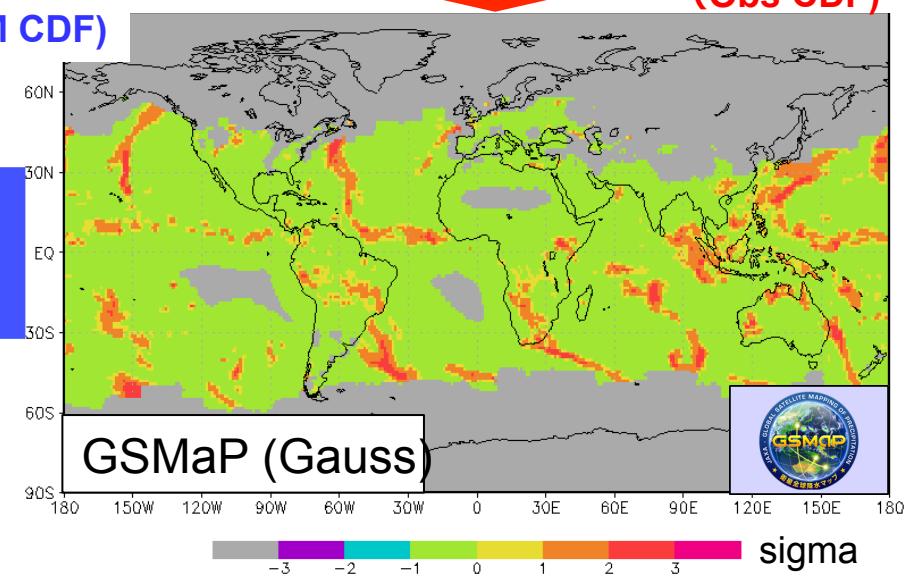
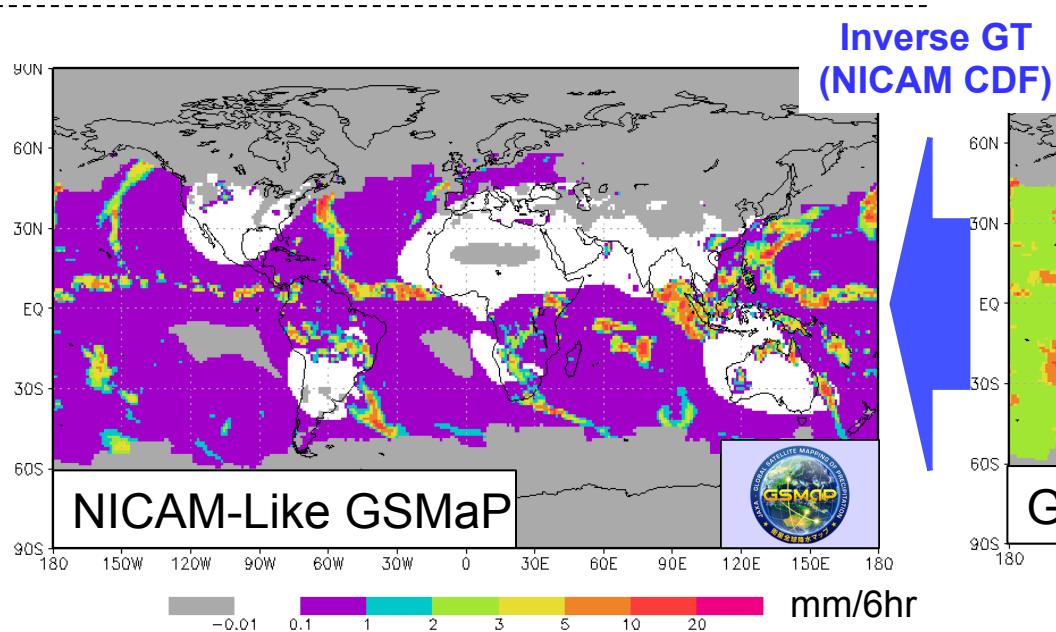
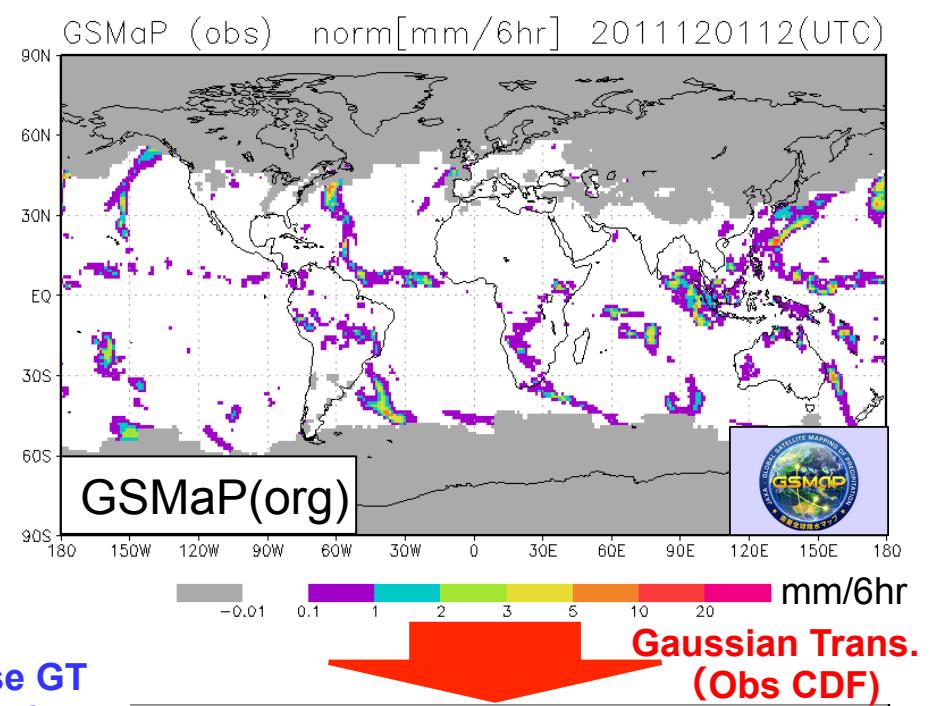
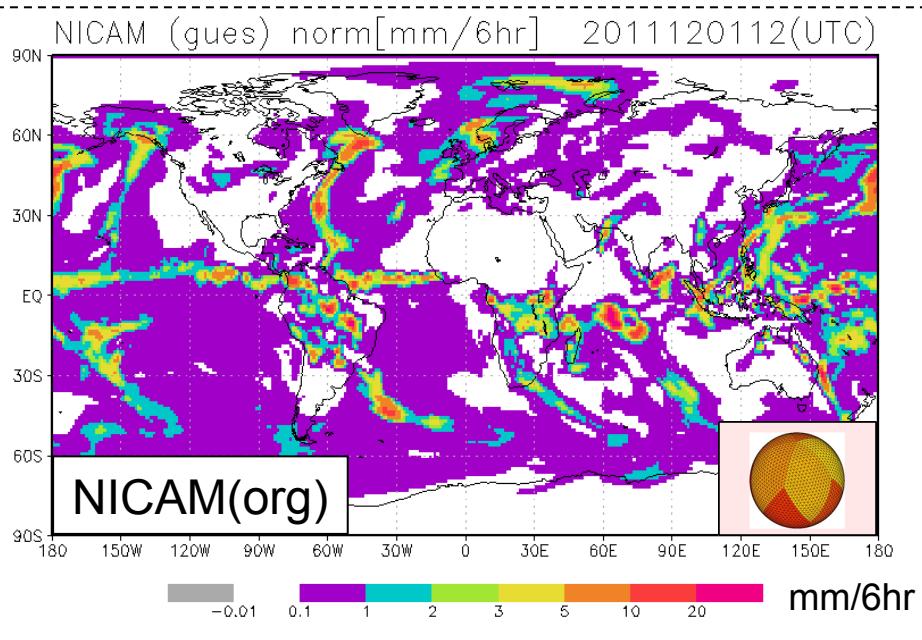


NICAM-like GSMAp using GT *(Kotsuki et al., 2015)*



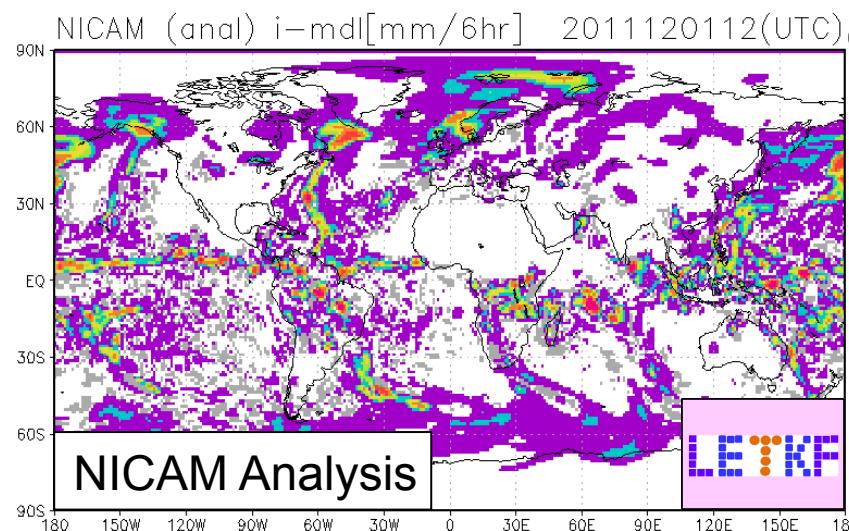
NICAM-like GSMAp using GT

(Kotsuki et al., 2015)



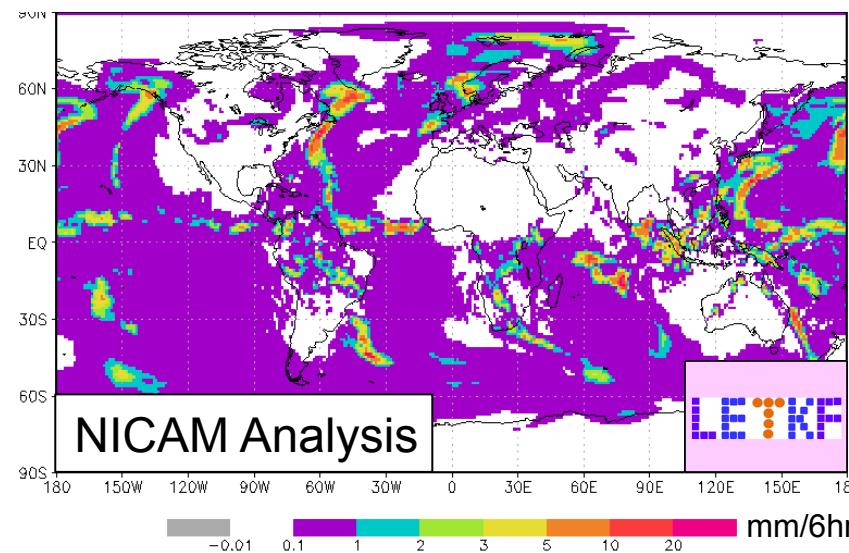
Successful attempt of precipitation DA using GT

w/o
GT



w/
GT

w/ Gaussian transformation

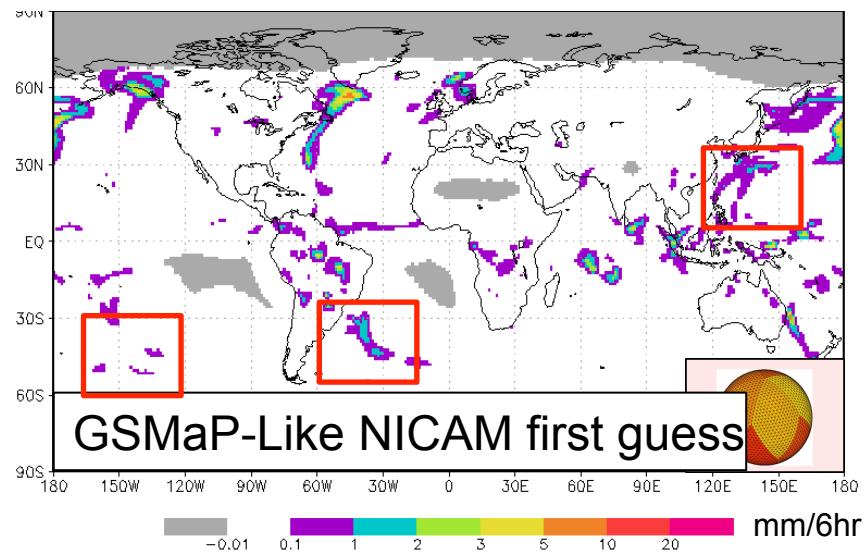
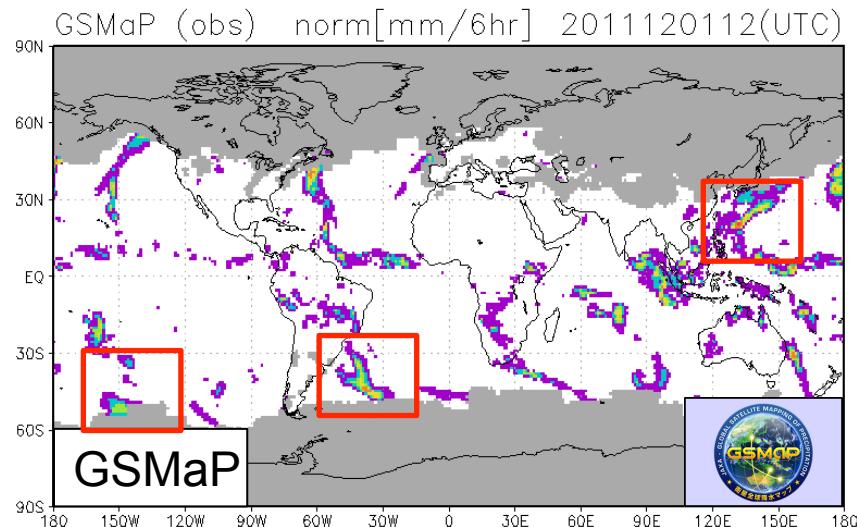


Successful attempt of precipitation DA using GT

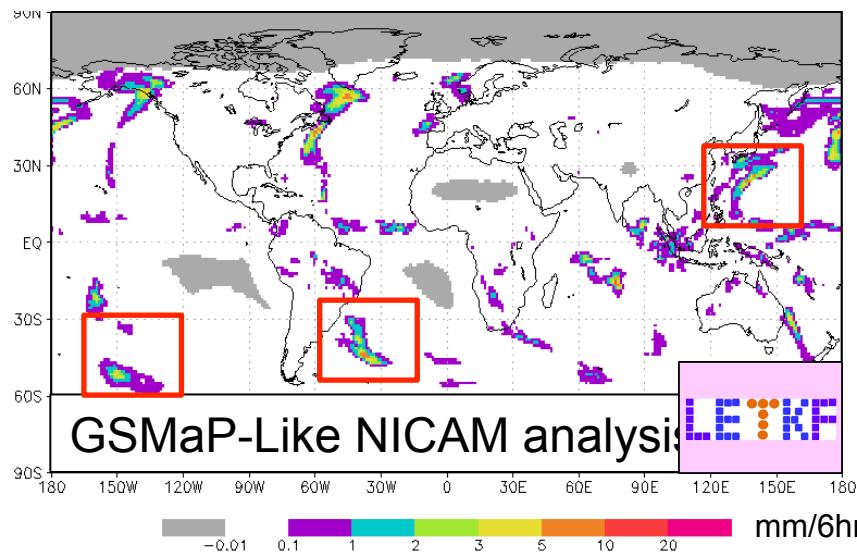
(Kotsuki et al., 2015)

Getting closer to obs due to DA

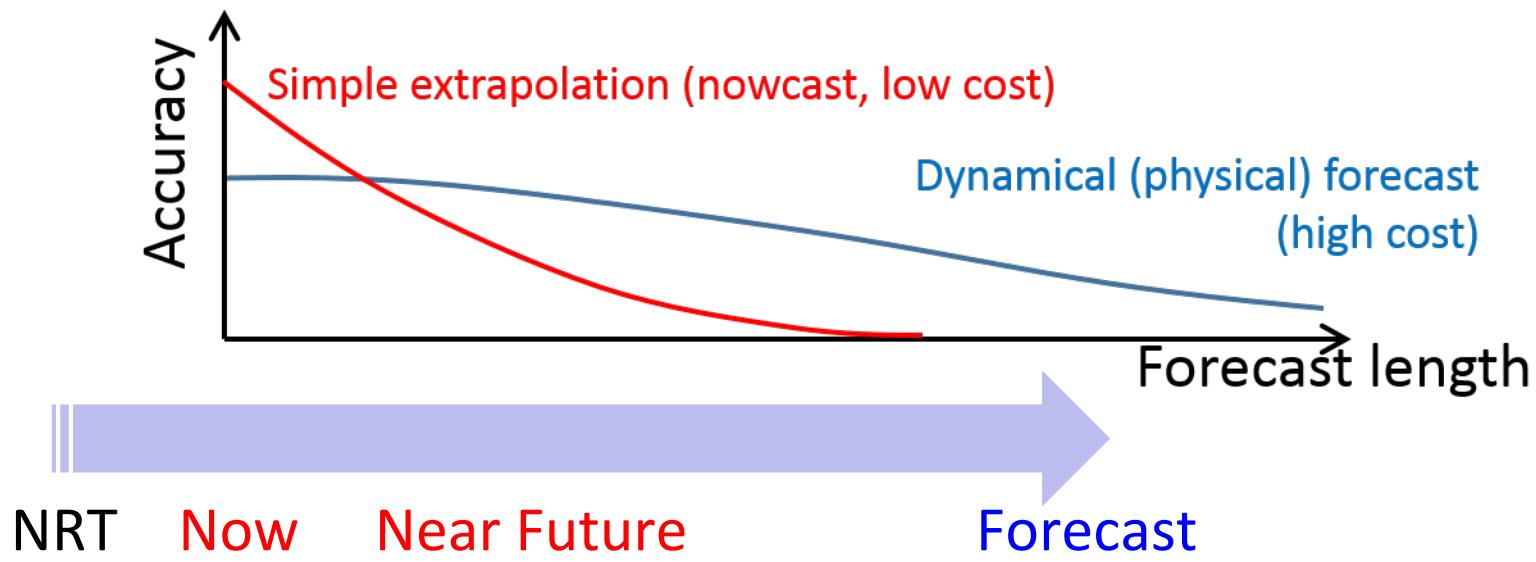
Analysis as satellite L4 product
(an idea suggested by M. Satoh)



DA



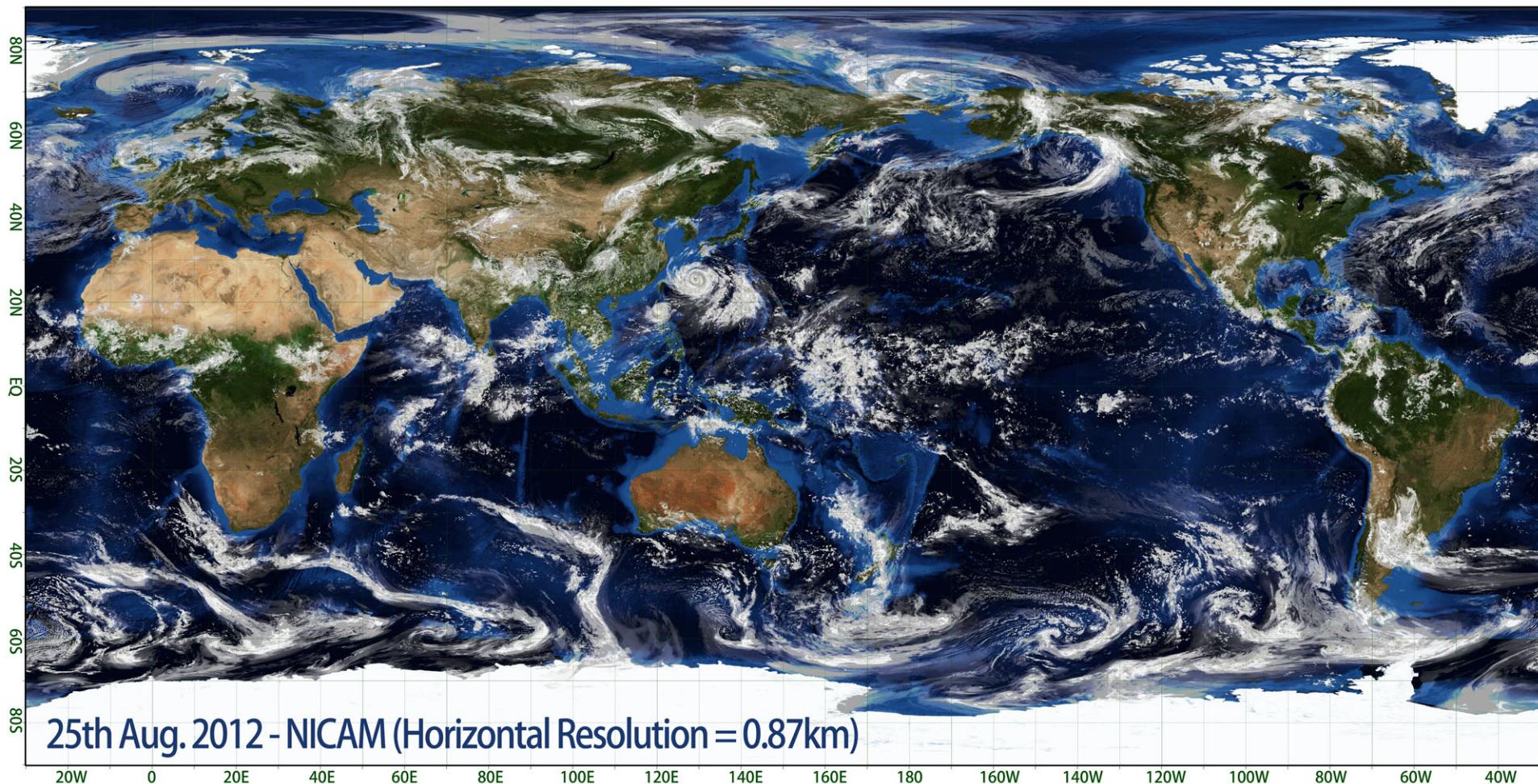
Seamless NRT → Now → NF → Forecast



A perspective

BIG SIMULATION

Global 870-m simulation (*Miyamoto et al. 2013*)



©JAMSTEC・AORI (SPIRE Field3), RIKEN/
AICS
Visualized by Ryuji Yoshida

Future perspectives

- “*Big Data Assimilation*” 30-sec. super-rapid DA cycle
 - Japanese “post-K” supercomputer planned in 2020
 - May “Tokyo 2020” be a good place to demonstrate?
- “*Big Data*” and “*Big Simulation*”
 - “Co-design” among modelers, observers, DA, and CS
- What would you do with such large computers?
 - Larger ensemble, higher resolution, multi-media coupling...

