Workshop on Sensitivity Analysis and Data Assimilation in Meteorology and Oceanography,
1-5 June 2015, Roanoke, WV, USA

Impact of initialization on the LETKF and observation thinning within JMA's global hybrid 4DVar-LETKF data assimilation system

Yoichiro Ota and Takashi Kadowaki

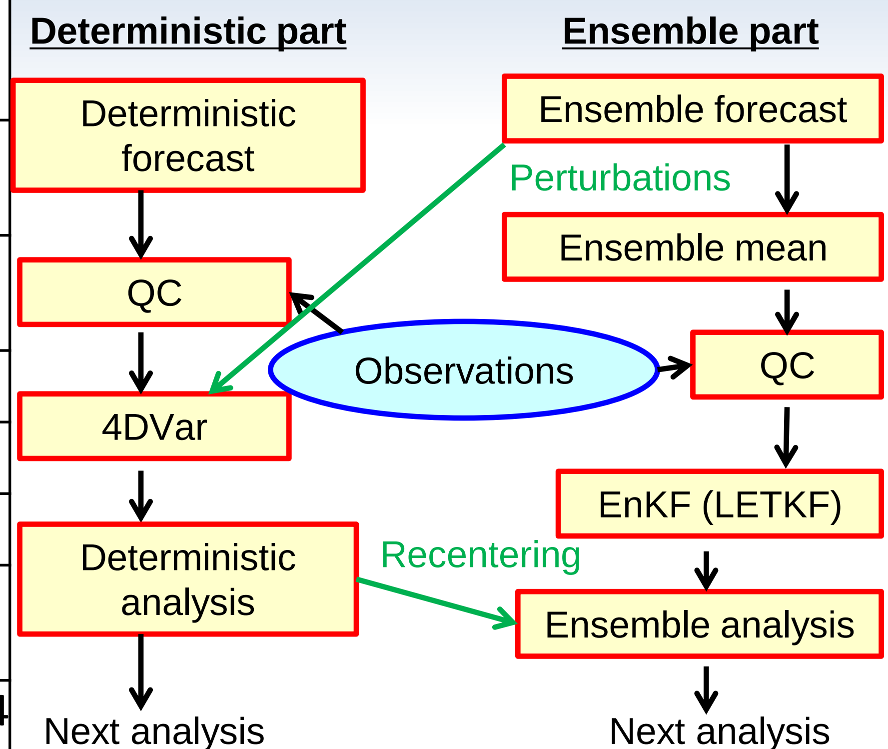
Japan Meteorological Agency / Numerical Prediction Division (JMA-NPD)

Outline

- Performance of the hybrid DA compared to the 4DVar in JMA operational global DA
- Recent developments
  - Introduction of the initialization using the surface pressure tendency analysis in the EnKF
  - Test of observation thinning in the LETKF to reduce computations and memory consumptions

Hybrid 4DVar-LETKF DA developed in JMA

Analysis resolution (outer / inner)	T_L959L100 (~20km, top:0.01hPa) / T_L319L100 (~55km, top:0.01hPa)
Assimilation window	6 hours (analysis time +/- 3 hours)
Hybrid method	Extended control variable method (Lorenc 2003)
Weights on $\mathbf{B}$	$\beta_{stat}{}^2 = 0.85$, $\beta_{ens}{}^2 = 0.25$
LETKF resolution	T_L319L100
Ensemble size	50
Localization scale (4DVar)	Horizontal: 800km Vertical: 0.8 scale heights
Localization scale (LETKF)	Horizontal: 400km, Vertical: 0.4 (0.8 for Ps) scale heights
Covariance inflation	Adaptive inflation (Miyoshi 2011)



$$\mathbf{B} = \beta_{stat}^2 \mathbf{B}_{stat} + \beta_{ens}^2 \mathbf{B}_{ens}$$

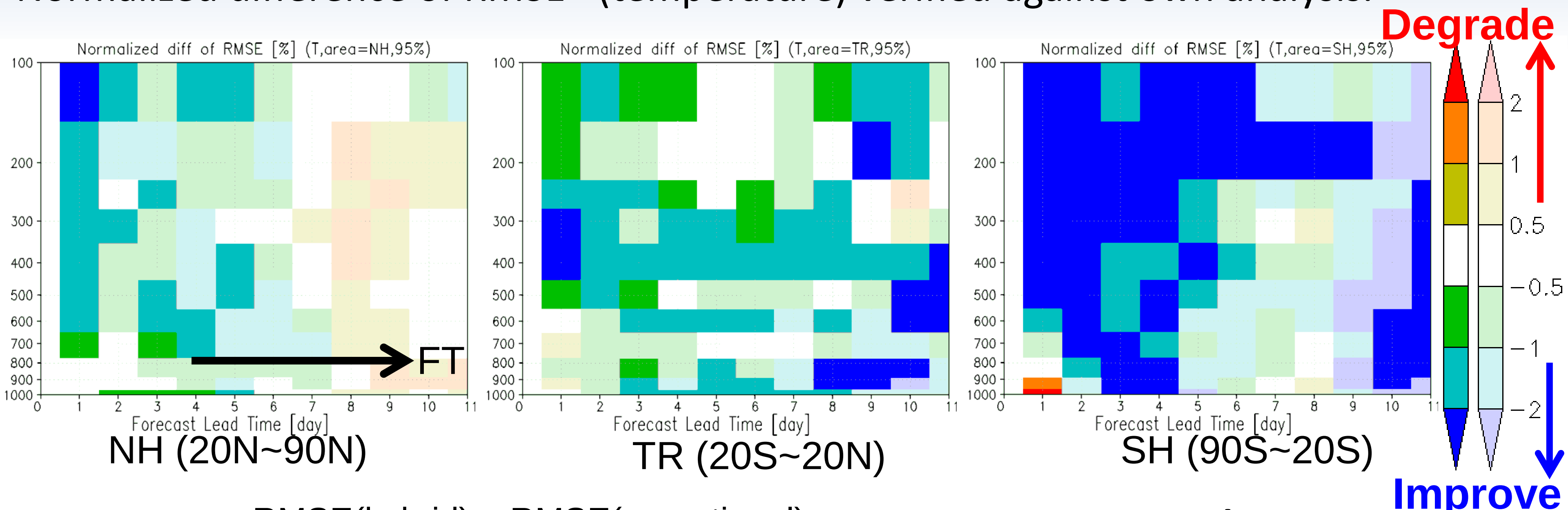
Static (Climatological) background error covariance

Ensemble-based background error covariance

Operational global DA at JMA is 4DVar (not hybrid)

Performance of hybrid DA compared to the operational system (1/2)

Normalized difference of RMSE* (temperature) verified against own analysis.



* defined as $\dfrac{\text{RMSE(hybrid)} - \text{RMSE(operational)}}{\text{RMSE(operational)}} \times 100\ \%$

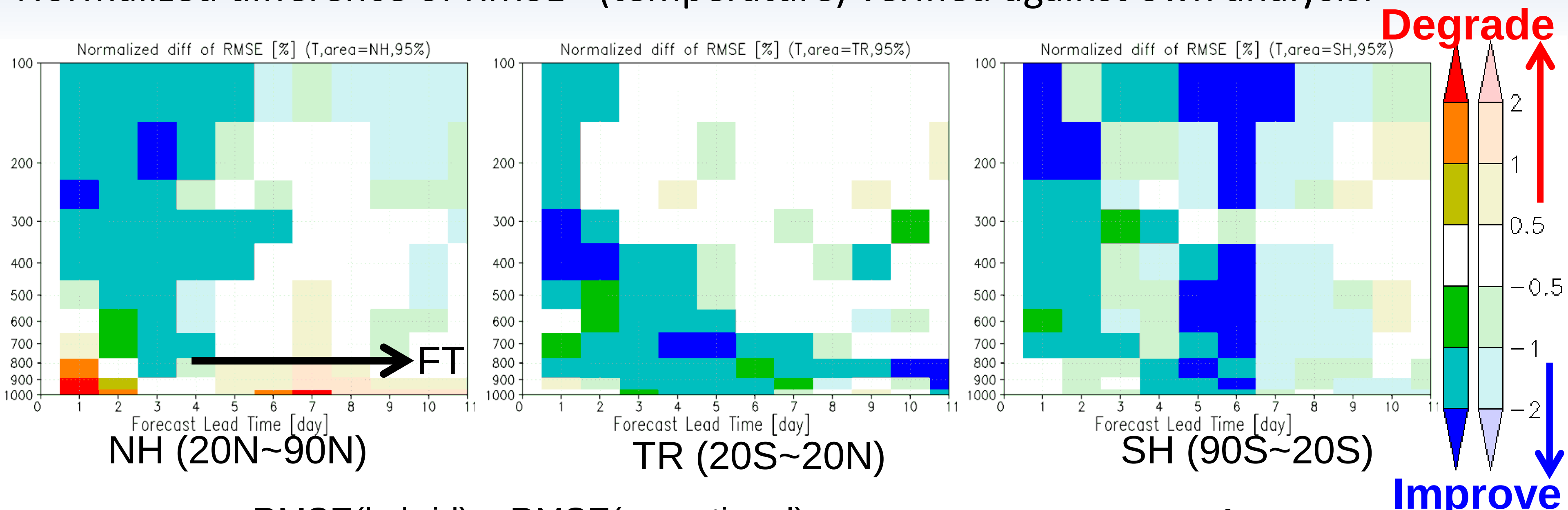
Intense colors
Statistically significant change at 95%

August 2013

Large forecast error reduction is found in the Southern Hemisphere extra-tropics (< -2%)

Performance of hybrid DA compared to the operational system (2/2)

Normalized difference of RMSE* (temperature) verified against own analysis.



* defined as $\frac{\text{RMSE(hybrid)} - \text{RMSE(operational)}}{\text{RMSE(operational)}} \times 100\ \%$

Intense colors
Statistically significant change at 95%

January 2014

The forecast error reduction is not as large as the boreal summer period (-1~-2%).

For further improvements (view points from ensemble part)

- To use ensemble perturbations that have better quality
  → Balanced ensemble analysis through the introduction of initialization
- To use larger number of ensemble members
  - Main obstacle is computational cost (i.e. CPU, memory, storage, etc…)
  - Future HPCs may have smaller memory per node.
    - It is crucial for memory intense jobs such as the LETKF.

  → Observation thinning

Outline

- Performance of the hybrid DA compared to the 4DVar in JMA operational global DA
- Recent developments
  - Introduction of the initialization using the surface pressure tendency analysis in the EnKF
  - Test of observation thinning in the LETKF to reduce computations and memory consumptions

Initialization on EnKF analysis (1/4)

Currently, no constraint is applied on the EnKF analysis for JMA/GSM.

- The EnKF analysis may contain the imbalances caused by the localization or the sampling errors with a limited number of ensemble members.
- Significant portion of the phase space spanned by the ensembles may be devoted to the non-growing mode such as the gravity waves.

To improve the balance of the perturbations generated from the LETKF and the hybrid DA, the initialization using surface pressure tendency analysis (Hamrud et al. 2014) has been tested.

- Test within the LETKF DA cycle (no recentering to the 4DVar analysis)
- Test within the Hybrid DA

Initialization on EnKF analysis (2/4)

Procedures

1) Add surface pressure tendency (dps/dt) to the state variables of the EnKF.
2) Based on the equation of continuity, dps/dt is equal to the convergence of the mass flux. Distribute the difference between analyzed and diagnosed dps/dt to the divergence of the upper atmospheric column.

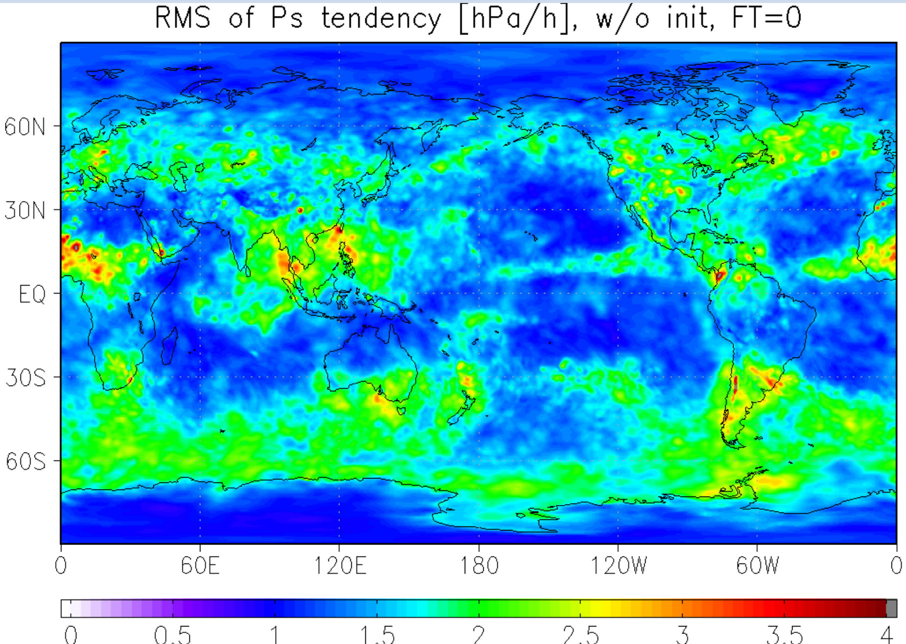
$$\frac{\partial p_s}{\partial t} = -\int_0^1 \nabla \cdot \left(\mathbf{v} \frac{\partial p}{\partial \eta} \right) d\eta \rightarrow \Delta\left[\nabla \cdot (\mathbf{v}_k dp_k)\right] = w_k \left[\left(\frac{\partial p_s}{\partial t} \right)_{diag} - \left(\frac{\partial p_s}{\partial t} \right)_{anl} \right] \quad w_k = \frac{S_{wind,k} dp_k}{\sum_{k=1}^{N} \left(S_{wind,k} dp_k \right)}$$

where p_s: surface pressure, $\mathbf{v}$: horizontal wind, S_{wind}: ensemble spread of wind, dp_k: thickness of the kth model layer

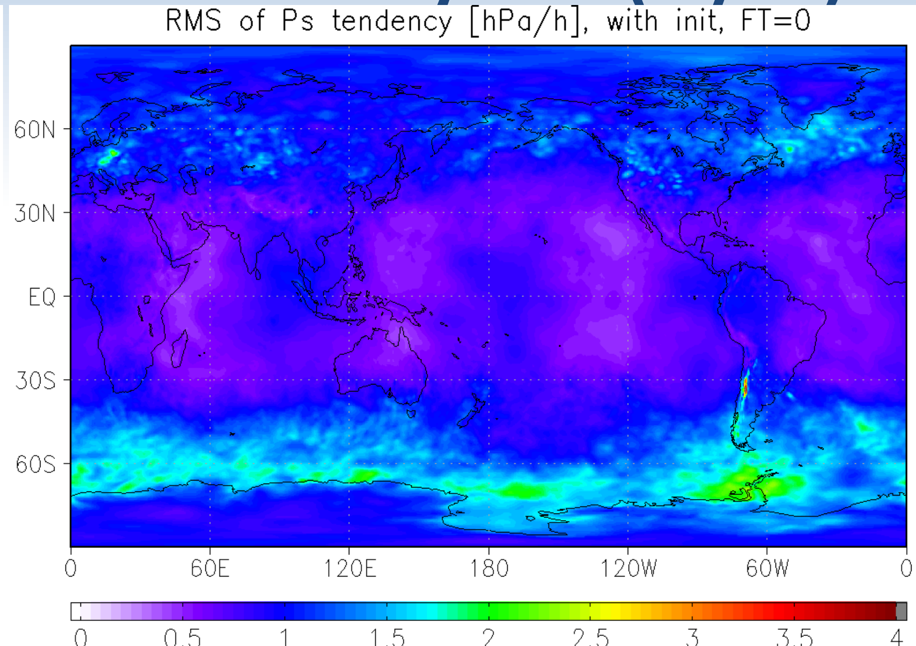
Weights w_k are determined by the analysis spread of the wind speed.

3) Assuming dp_k is constant before and after the initialization, the wind increments are derived.

Initialization on EnKF analysis (3/4)



without initialization



with initialization

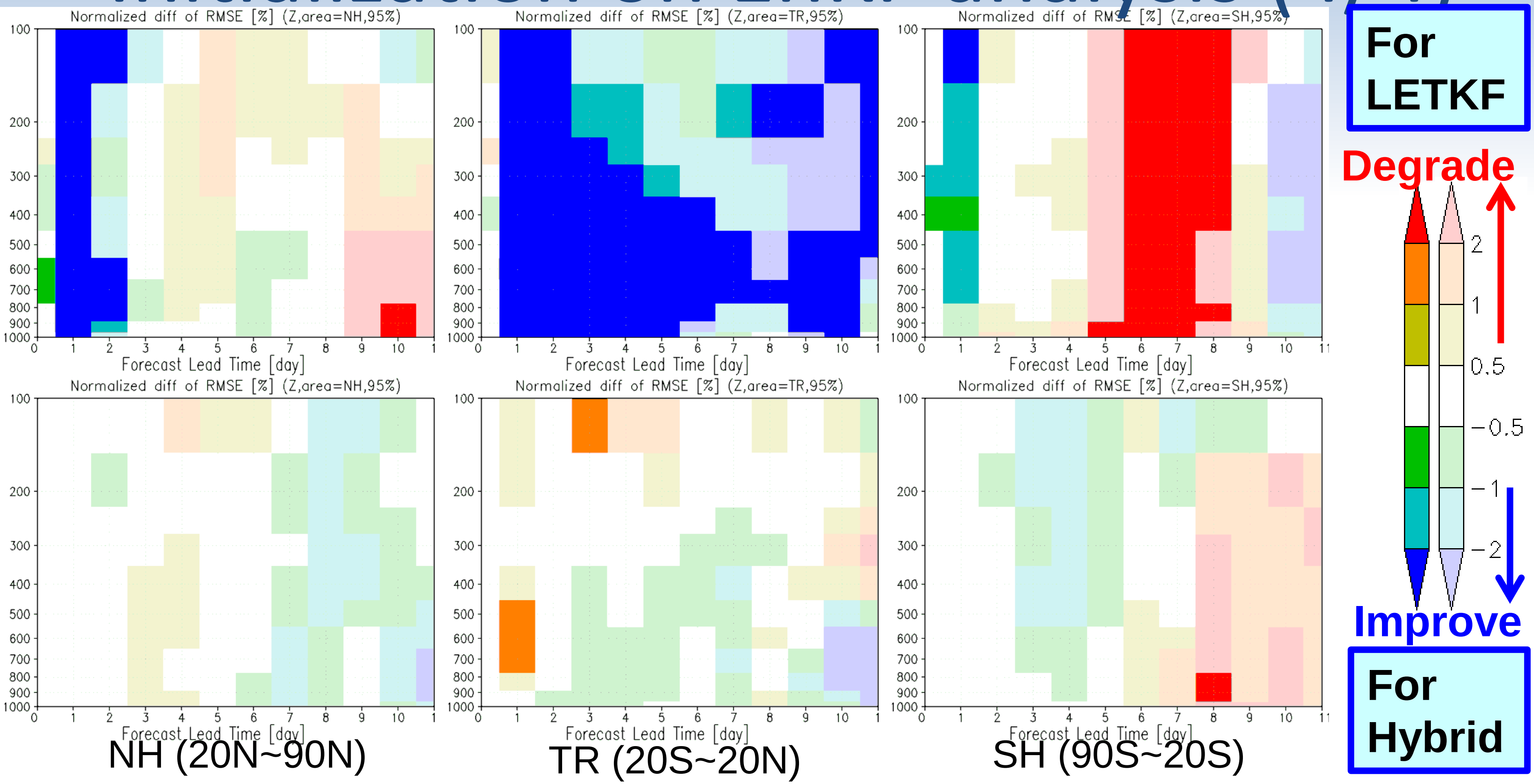
RMS of the surface pressure tendency at FT=0 of forecast ensemble initiated from the LETKF analysis (hPa / h, August 2013)

Excessive gravity waves can be reduced by the initialization.

	NH(20N-90N)		TR (20S-20N)		SH (90S-20S)	
	O-A	O-B	O-A	O-B	O-A	O-B
w/o initialization	0.601	0.887	0.636	0.937	0.825	1.134
with initialization	0.601	0.862	0.635	0.918	0.826	1.115

RMS of O-A and O-B of the SYNOP surface pressure observations (hPa) for the LETKF cycle experiment.

Initialization on EnKF analysis (4/4)



Normalized difference of RMSE (geopotential height) verified against own analysis with and without initialization.

Note: August 2013, AIRS and IASI are not assimilated

- For the LETKF, short-range forecast RMSE is significantly decreased.
- For the hybrid, the impact on the forecast accuracies is not as large as for the LETKF.

Outline

- Performance of the hybrid DA compared to the 4DVar in JMA operational global DA
- Recent developments
  - ➢ Introduction of the initialization using the surface pressure tendency analysis in the EnKF
  - ➢ Test of observation thinning in the LETKF to reduce computations and memory consumptions

Strategies on reducing the computational cost

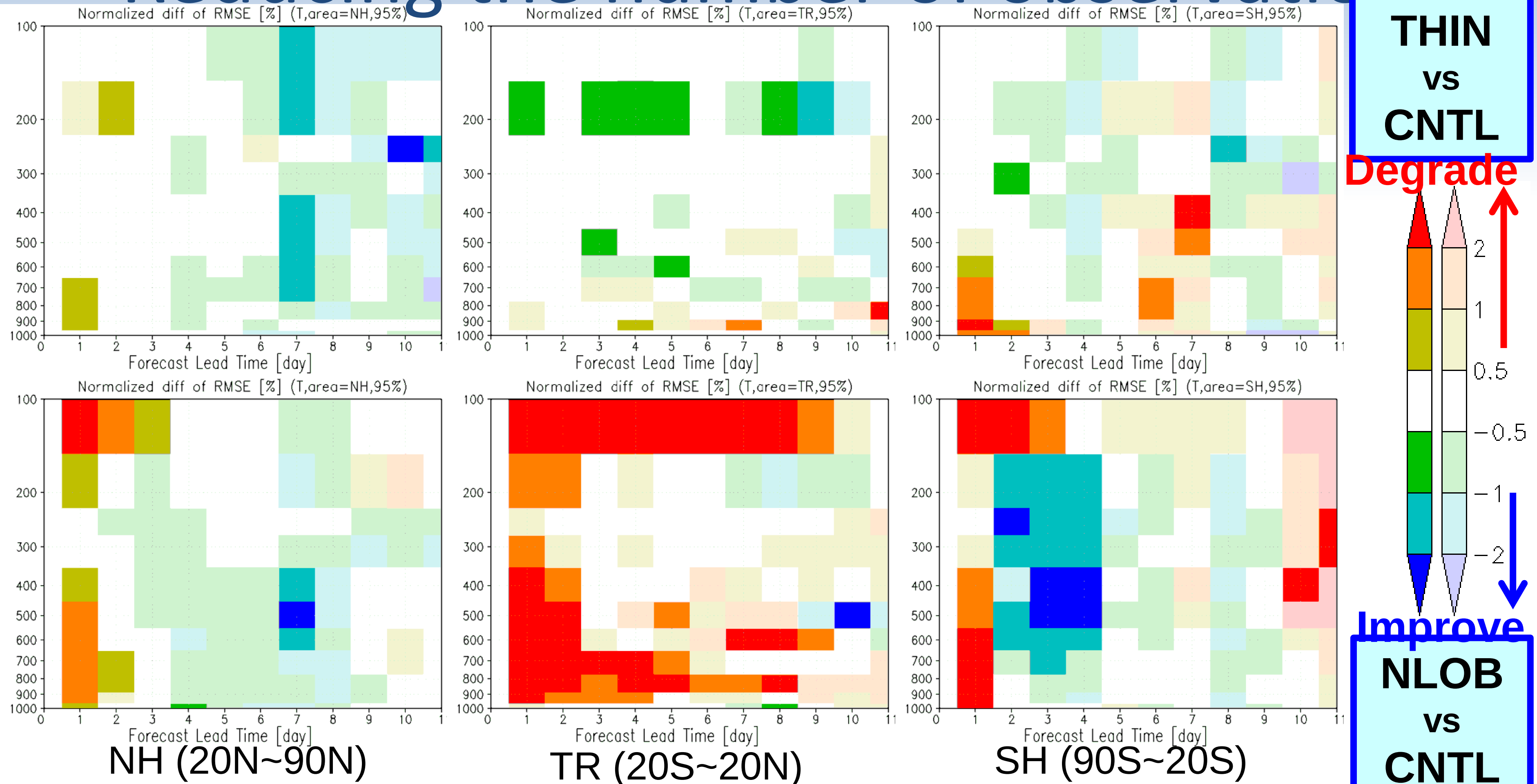
- LETKF requires large memory when the large number of observations and/or ensemble members are used.
- EnKF can extract a limited amount of information from the large number of observations using small ensembles.
  - Currently, number of observations assimilated on local (O10^3-10^4) >> number of ensemble members (O10-10^2)
- In the hybrid DA, the number of observations assimilated in the EnKF may be reduced w/o significant degrade on the analysis quality.
  - This will reduce the memory consumption and computations.
  - Some earlier studies (e.g. Migliorini 2013, Hamrud et al 2014) suggest that limiting the number of observations assimilated on local has neutral (or even beneficial) impact on the EnKF analysis.

We tested observation thinning for the LETKF in the hybrid DA

Experimental settings

- In CNTL, the LETKF assimilates as much observations as the 4DVar.
- In THIN, the LETKF assimilates about an half of observations used in the 4DVar
  - Observation thinning interval is increased for the LETKF.
- In NLOB, the LETKF assimilates 50 observations at maximum on local. Observations are selected, based on the value of the localization function (closest set to the analysis grid).
  - The number is the same as the number of ensemble members. ← Based on the fact that the local dimension spanned by the ensemble is N_{mem}-1 at most.
  - Computational time is almost the same as THIN.

Reducing the number of observations



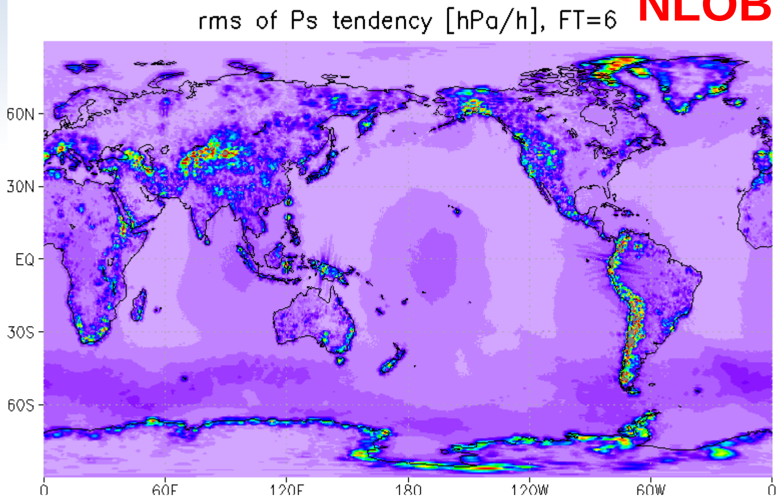
Normalized difference of RMSE (temperature) verified against own analysis.

Note: August 2013

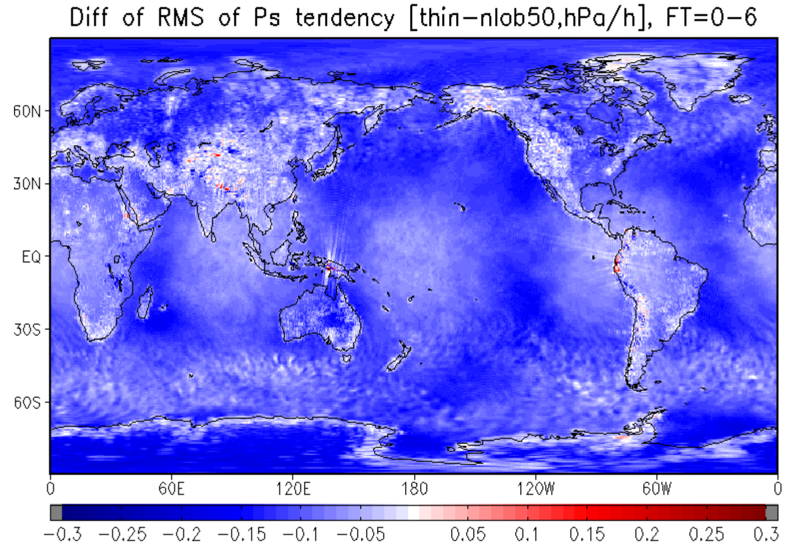
- For THIN, the impact on the forecast RMSE is almost neutral.
- For NLOB, the forecast RMSE is significantly increased especially in the Tropics.

Comparison of Surface pressure tendency

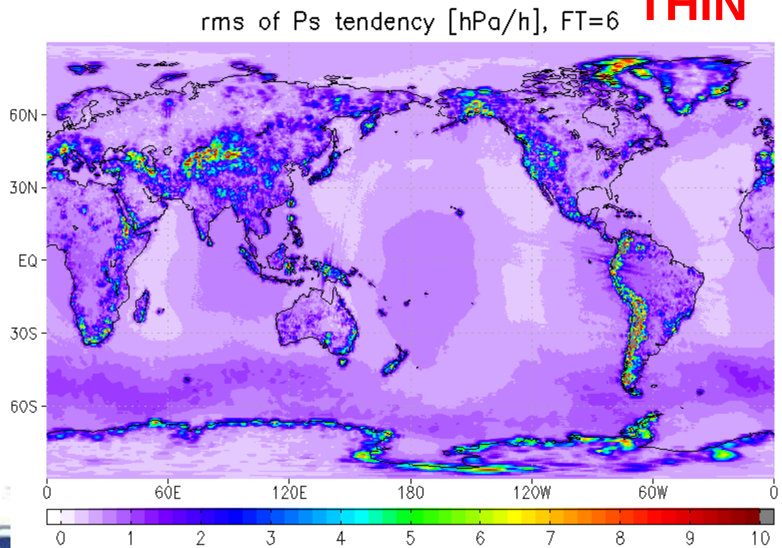
NLOB



Diff THIN-NLOB



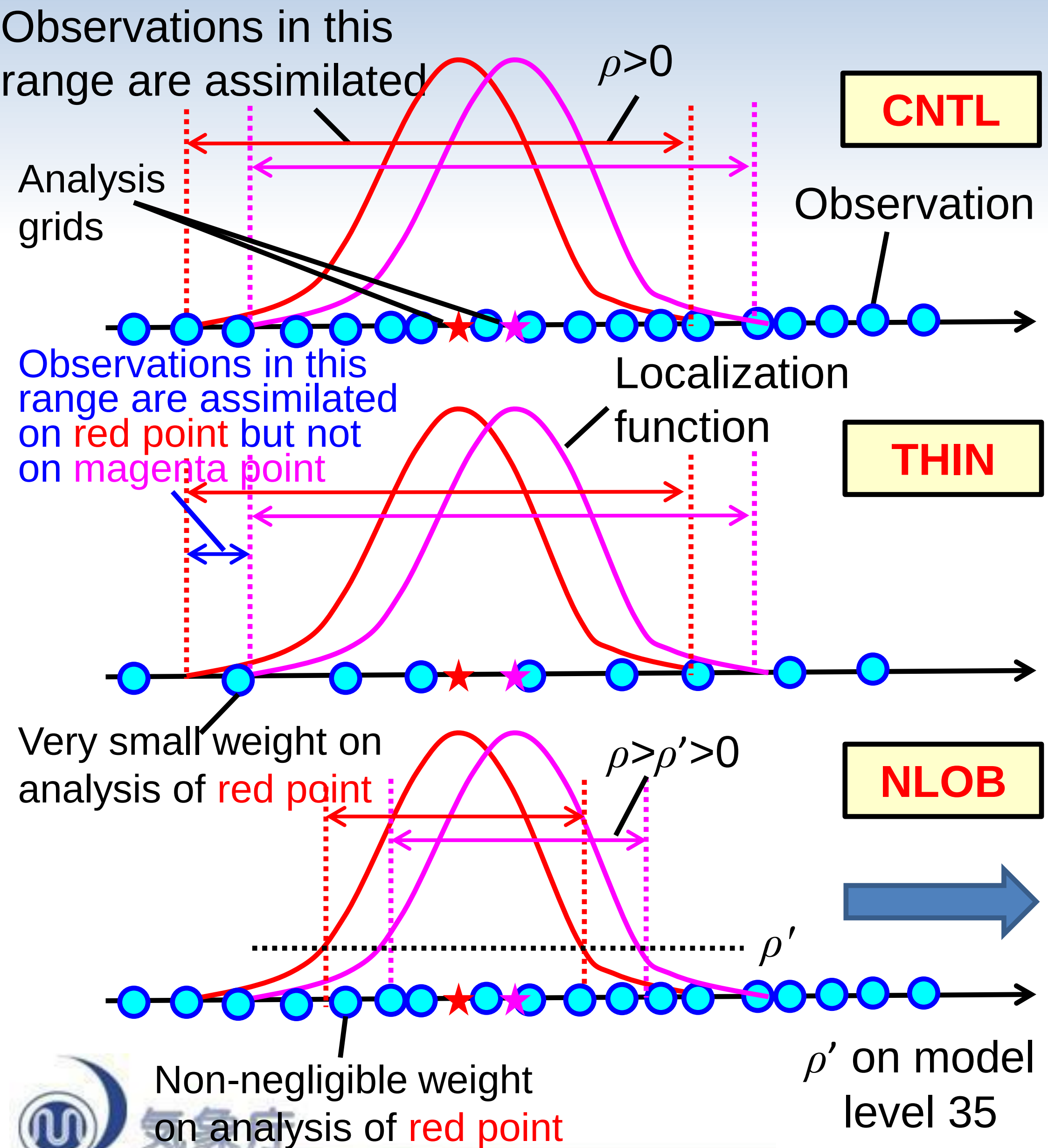
THIN



Comparison of RMS of the surface pressure tendency in ensemble forecasts (FT=0-6)

- Large tendency near steep orography
  - Due to the resolution change
- Tidal waves (wavenumber 4)
- Larger tendency for NLOB compared to that of THIN
  → suggesting that NLOB introduces imbalance to the LETKF analysis

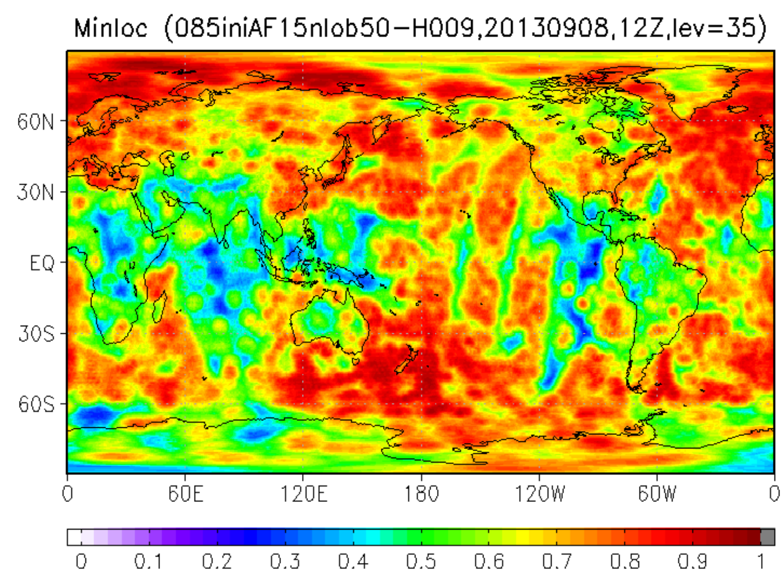
Possible cause of imbalance in NLOB



For NLOB,

- Localization scale is effectively smaller by limiting the observation use.
- Observations assimilated on local can be much different from those of adjacent grid
  - These observations may have non-negligible weights on the analysis

This may introduces large imbalance on the analysis.



Japan Meteorological Agency

Better strategies?

- Super-observations
  - Information loss is supposed to be smaller than the simple observation thinning.
- Select observations based on ensemble DFS (Degrees of Freedom for Signal, Liu et al. 2009)
  - Only needs R and HXa which can be derived before the LETKF update on analysis grid.

$$\mathbf{S}_o = \mathbf{R}^{-1}\mathbf{H}\mathbf{P}_a\mathbf{H}^{\mathrm{T}} = \frac{1}{K-1}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{X}_a\right)\left(\mathbf{H}\mathbf{X}_a\right)^{\mathrm{T}}$$

We will continue the investigations for further refinements.

Summary

- JMA is developing global hybrid 4DVar-LETKF as a possible candidate for future DA system.
- Initialization on the EnKF ensembles using surface pressure tendency analysis has been tested.
  - It can reduce the imbalance in the EnKF analysis.
  - The impact is smaller when it is applied to the hybrid DA.
- Simple observation thinning in the EnKF analysis can be applied without significant degrade on the quality of hybrid DA analysis.
  - It can reduce computations and memory consumptions
    → It will allow the use of more ensemble members
  - More sophisticated strategies may be applied for future developments

Backup slides

Hybrid 4DVar with extended control variables

Using the extended control variable α which is defined as the weight given to the each ensemble member, the cost function of the hybrid 4DVar-LETKF (based on Lorenc 2003 and Buehner 2005) is expressed as

$$J(\mathbf{x}',\alpha_1,\ldots,\alpha_k)=\frac{1}{2}\mathbf{x}'^{\mathbf{T}}\mathbf{x}'+\frac{1}{2}\sum_{k=1}^{K}\alpha_k^{\mathbf{T}}\alpha_k$$

$$+\frac{1}{2}\left[\mathbf{HM}\left(\beta_1\mathbf{B}^{1/2}\mathbf{x}'+\beta_2\sum_{k=1}^{K}\mathbf{X}'_k\circ\left(\mathbf{C}^{1/2}\alpha_k\right)\right)-\mathbf{y}'\right]^{\mathbf{T}}\mathbf{R}^{-1}\left[\mathbf{HM}\left(\beta_1\mathbf{B}^{1/2}\mathbf{x}'+\beta_2\sum_{k=1}^{K}\mathbf{X}'_k\circ\left(\mathbf{C}^{1/2}\alpha_k\right)\right)-\mathbf{y}'\right]+J_c$$

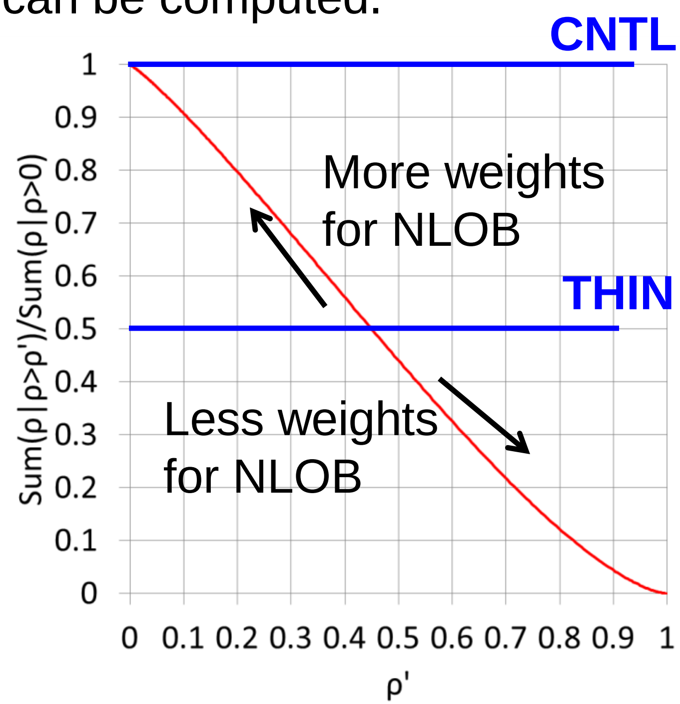
weights on static and ensemble-based B

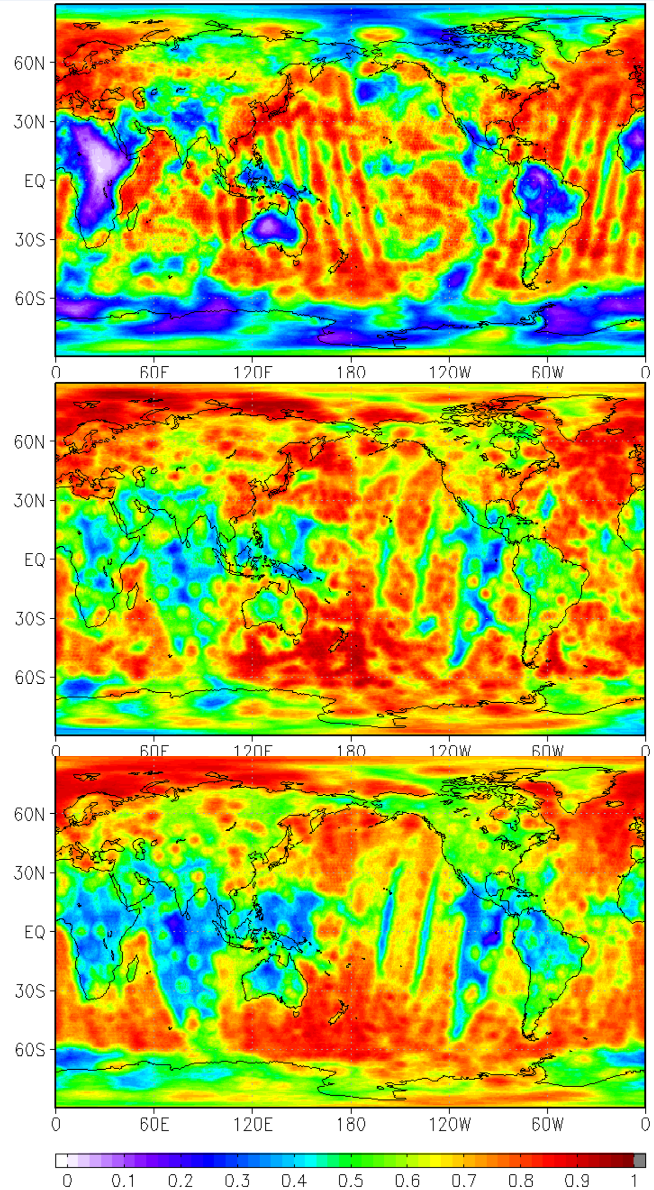
$\beta_1 = 0.85$, $\beta_2 = 0.25$ are used for the experiments

B: (static) background error covariance, **H**: observation operator, **R**: observation error covariance, **M**: tangent linear model, **y′**: innovation, **x′**: control variables, α: extended control variable, **X′**: ensemble perturbations, **C**: localization function

Comparison of "weights" to observations in THIN and NLOB

Assuming that the observations are uniformly distributed, summation of localization function larger than ρ' can be computed.



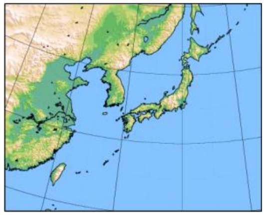


ρ' on the lowest model layer

ρ' around 450 hPa

ρ' around 130 hPa

Operational Deterministic NWP Models at JMA

	Global Spectral Model (GSM)	Meso-Scale Model (MSM)	Local Forecast Model (LFM)
Purposes	Short- and medium-range forecast	Warnings and very short-range forecast	Disaster prevention and Aviation forecast
Forecast domain	Global	Japan and its surrounding areas 	Japan and its surrounding areas 
Horizontal resolution	Approximately 20km (TL959)	5km/ 817x661	2km/ 1581x1301
Vertical levels/ Top	100 / 0.01hPa	50 / 21800m	60 / 20200m
Forecast hours (initial time)	84 hours (00, 06, 18 UTC), 264 hours (12 UTC)	39 hours (00, 03, 06, 09, 12, 15, 18, 21 UTC)	9 hours (00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 UTC)
Analysis	4D-Var	4D-Var	3D-Var

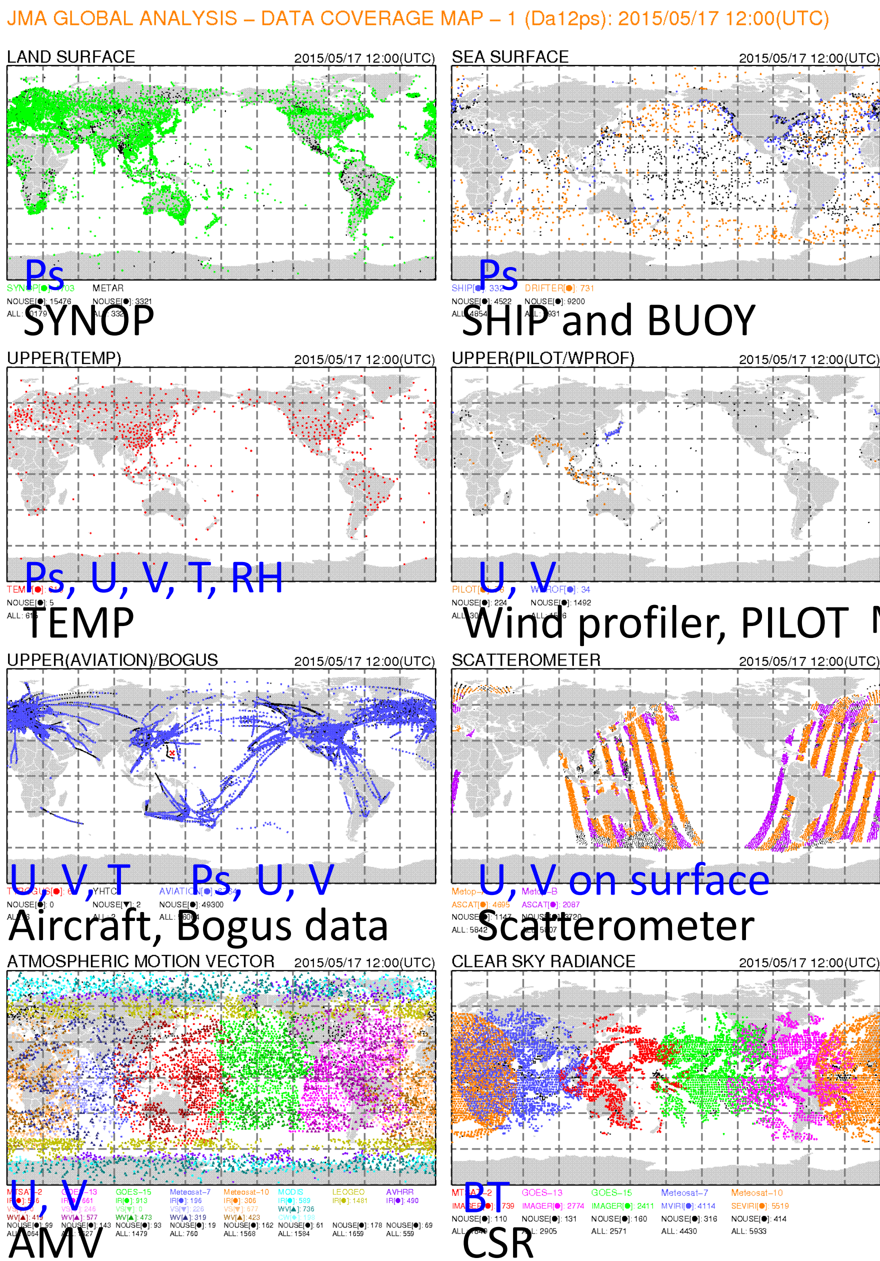
Operational Global Analysis

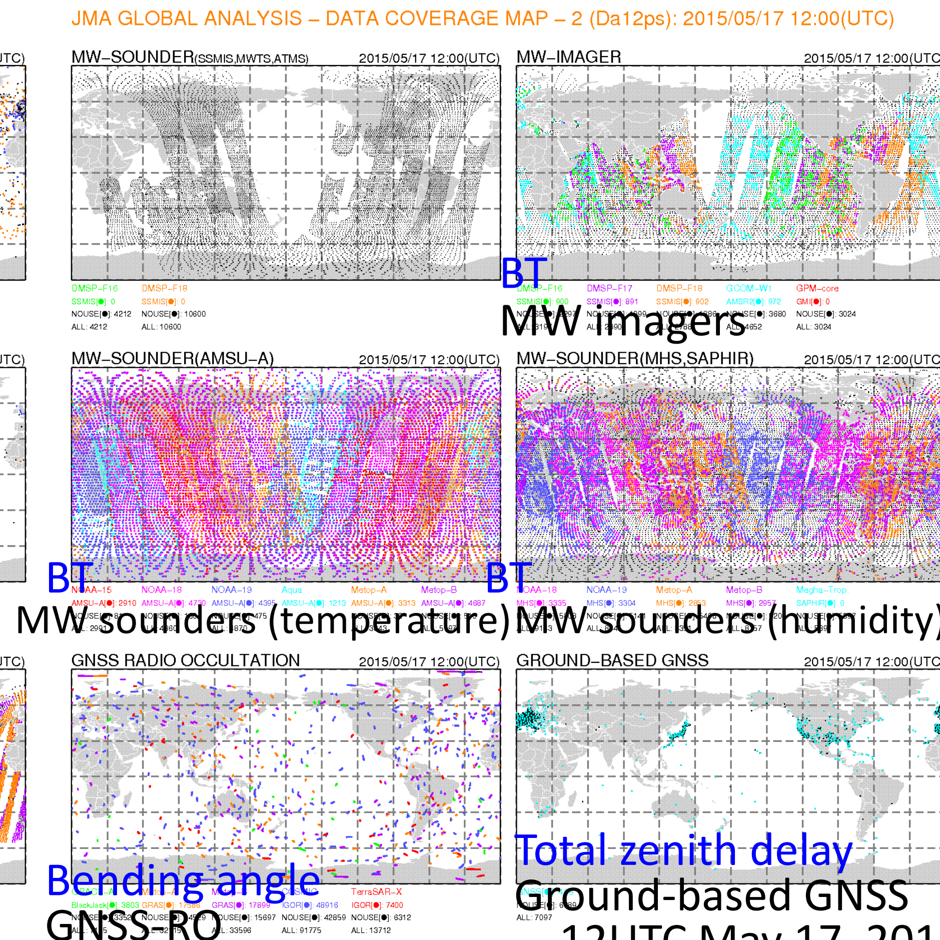
Cut-off time	2h20m for early run analyses at 00, 06, 12 and 18 UTC, 11h50m for cycle run analyses at 00 and 12 UTC, 7h50m for cycle run analyses at 06 and 18 UTC
Initial Guess	6-hour forecast by GSM
Grid form, Horizontal resolution	Reduced Gaussian grid, approximately 20km for outer model Reduced Gaussian grid, approximately 55km for inner model
Vertical resolution	100 forecast model levels up to 0.01 hPa + surface
Analysis variables	Surface pressure, temperature, winds and specific humidity
Methodology	Four-dimensional variational (4D-Var) scheme on model levels
Data Used	SYNOP, SHIP, BUOY, TEMP, PILOT, wind profiler, Aircraft, MW sounders, MW imagers, Hyper-spectral IR sounders, sea surface wind data from scatterometer on the Metop, AMV and CSR from geostationary satellites, MODIS wind data from Terra and Aqua, bending angle from GNSS radio occultation observation, GNSS total zenith delay; Typhoon bogussing applied for analysis
Initialization	Non-linear normal mode initialization and a vertical mode initialization for inner model

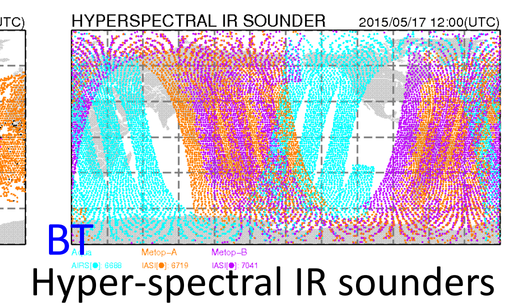
Early Analysis: Analysis for weather forecast. The data cut off time is very short.

Cycle Analysis: Analysis for keeping quality of global data assimilation system. This analysis is done after much observation data are received.

Observations assimilated in global analysis







12UTC May 17, 2015

Blue: assimilated elements

(Ps: surface pressure, U,V: horizontal winds, T: temperature, RH: relative humidity, BT: brightness temperature)