Comparison between LETKF and EnVAR with observation localization

*[1]Sho Yokota, [1]Masaru Kunii, [1]Kazumasa Aonashi, [1]Seiji Origuchi, [2,1]Le Duc, [1]Takuya Kawabata, [3,1]Tadashi Tsuyuki

[1]Meteorological Research Institute, [2]JAMSTEC, [3]Meteorological College

2016.2.17 Data Assimilation Seminar in RIKEN/AICS

Ensemble-based variational data assimilation

Bayesian assimilation provides Analysis $\mathbf{x}_0$ from First guess $\mathbf{x}_0^f$ and Observation $\mathbf{y}_t$.
$\mathbf{x}_0$ takes a maximum likelihood value when cost function J is minimum ($\nabla J=0$).

Background term | Observation term

Cost Function

$$J = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^f\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^f\right) + \frac{1}{2}\sum_t \left[H\left(M_t(\mathbf{x}_0)\right) - \mathbf{y}_t\right]^T \mathbf{R}_t^{-1}\left[H\left(M_t(\mathbf{x}_0)\right) - \mathbf{y}_t\right]$$

Gradient

$$\nabla J \equiv \frac{\partial J}{\partial \mathbf{x}_0} = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^f\right) + \sum_t \left[\boxed{\frac{\partial H\left(M_t(\mathbf{x}_0)\right)}{\partial \mathbf{x}_0}}\right]^T \mathbf{R}_t^{-1}\left[H\left(M_t(\mathbf{x}_0)\right) - \mathbf{y}_t\right]$$

Several methods are classified using how to solve $\nabla J=0$.

	Background covariance	How to solve x_0	How to calculate
3DVAR,4DVAR	Statistic	Implicitly	With adjoint of M and H
EnKF	Ensemble-based	Explicitly	Ensemble approximation
Hybrid-4DVAR	Ensemble-based	Implicitly	With adjoint of M and H
EnVAR	Ensemble-based	Implicitly	Ensemble approximation

EnVAR provides analysis implicitly without adjoint models

Why is ∇J=0 solved implicitly?

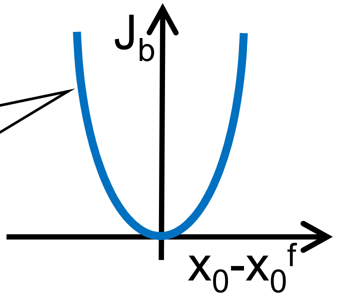
Background term

Observation term

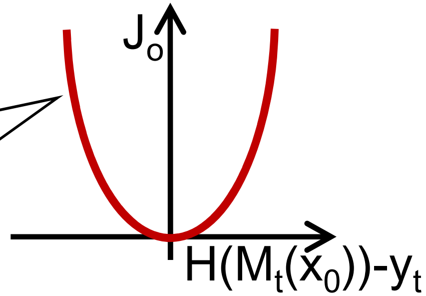
Cost Function

$$J=\underbrace{\frac{1}{2}(\mathbf{x}_0-\mathbf{x}_0^f)^T\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}_0^f)}_{\text{Background term}}+\underbrace{\frac{1}{2}\sum_t[H(M_t(\mathbf{x}_0))-\mathbf{y}_t]^T\mathbf{R}_t^{-1}[H(M_t(\mathbf{x}_0))-\mathbf{y}_t]}_{\text{Observation term}}$$

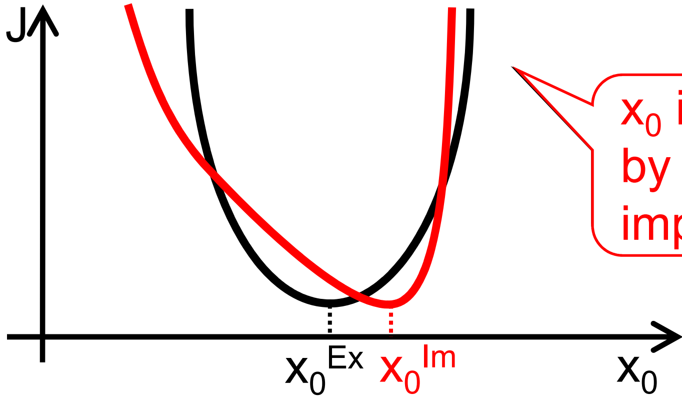
Quadratic function of x_0



Quadratic function of $H(M_t(x_0))$



$H(M_t(x_0))$ is strictly treated

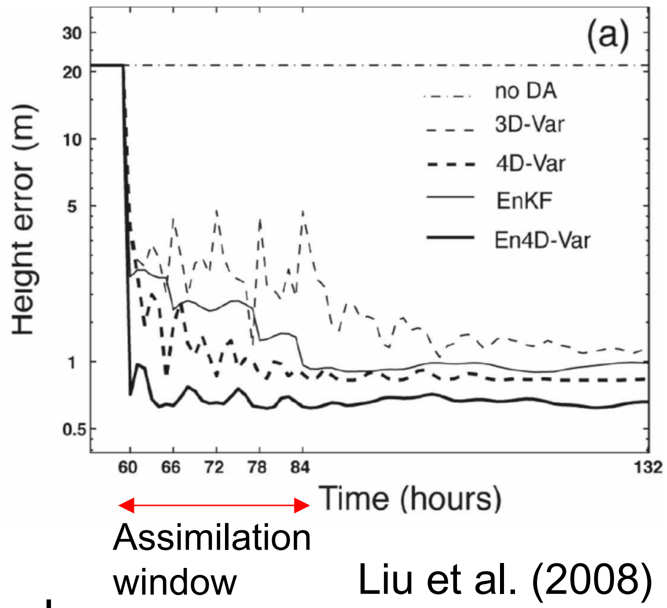


x_0 is changed by solving ∇J=0 implicitly

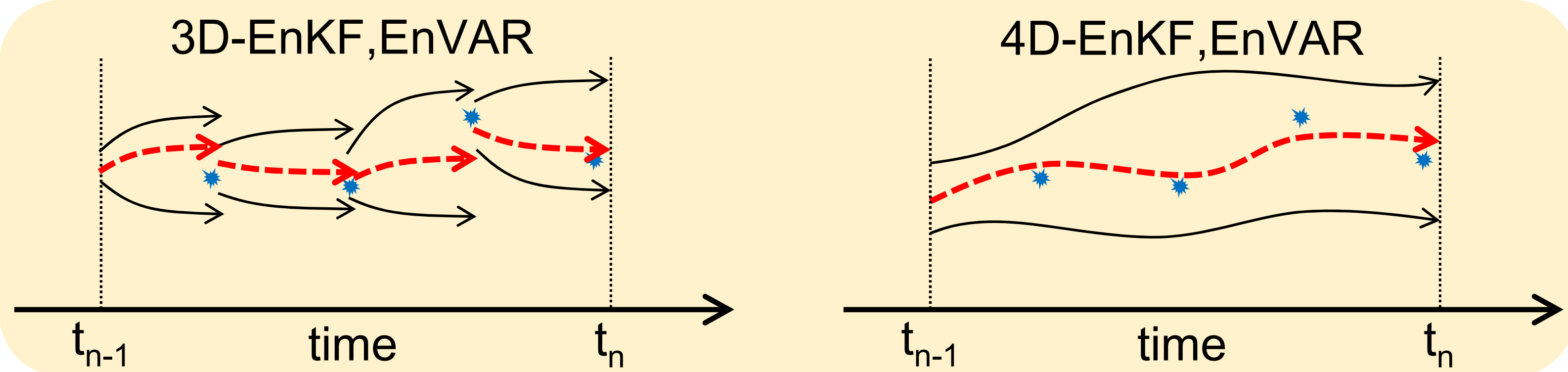
Implicit method is better because observation operator are strictly treated

Previous studies about EnVAR

- Zupanski (2005), Zupanski et al. (2008)
  - EnVAR method is suggested
- Liu et al. (2008, 2009), Buehner (2013)
  - 4D-EnVAR is compared to other methods
  - Gaussian approximation
  - Spatially localized background covariance
- Hunt et al. (2004)
  - 4D-EnKF is suggested
  - Any time analysis in assimilation window is provided



Liu et al. (2008)

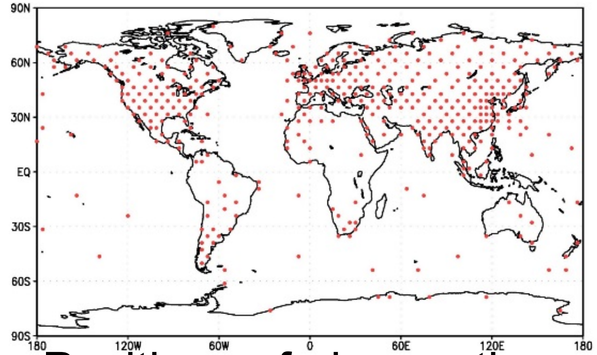


In this study, 4D-EnVAR is compared to 4D-EnKF (LETKF)

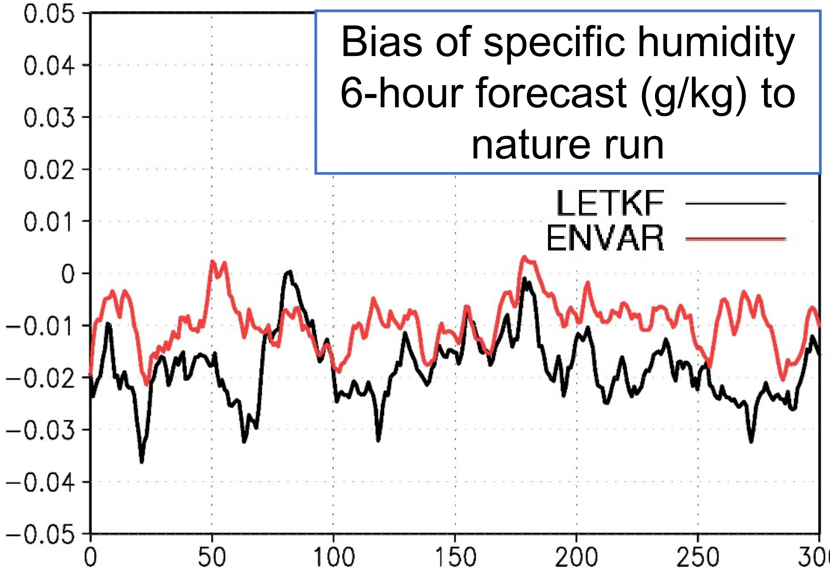
Contents

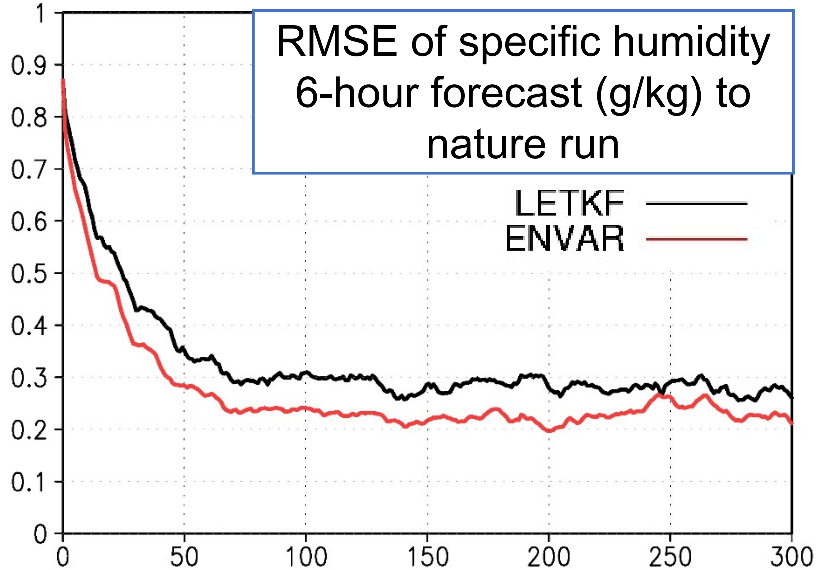
Observation system simulation experiments with SPEEDY model

Number of members: **20**
Assimilation window: **6 hours**
Localization radius: **σ_H=1000(km), σ_V=0.1(sigma)**
Observations: **U, V, T, RH, Ps**
Inflation: **Multiplicative (=1.1)**
Analysis time: **Center of the window (t=3h)**
Observation time: **t=1h, 3h, 5h**



Positions of observations





Number of forecast-analysis cycles (every 6 hour)

Bias and RMSE of EnVAR are smaller than those of LETKF

Formulation

EnVAR formulation

Multiplicative inflation

Analysis

$$\overline{\mathbf{x}_t^a} = \overline{\mathbf{x}_t^f} + \sum_j \delta\mathbf{x}_{j,t}^f \circ \mathbf{w}_j$$

Analysis valuable is transformed from **x** to **w**

Background error covariance

$$\mathbf{B} = \frac{\alpha^2}{M-1}\mathbf{L} \circ \sum_j \delta\mathbf{x}_{j,0}^f \delta\mathbf{x}_{j,0}^{f\ T}$$

Localization

i: analysis points
k: observation points
t: time slots
j: ensemble members

Then,

Gradient of J

Perturbation of H

$$\frac{\partial J}{\partial w_{ij}} = \frac{M-1}{\alpha^2} w_{ij} + \sum_{k,t} \frac{1}{R_{k,t}} L_{ik} \delta H_{kj,t} \left[H_k\left(\overline{\mathbf{x}_t^a}\right) - y_{k,t} \right]$$

Observation error variance

Observation operator (non-linear)

- **Observation localization**
- **Derived from Background localization**
- **Globally defined J**

Similar to LETKF (Observation localization)

Different from LETKF (Derived from Background localization; Globally defined J)

EnVAR without localization

Cost function: $J=\frac{1}{2}\left(\overline{\mathbf{x}_0^a}-\overline{\mathbf{x}_0^f}\right)^T\mathbf{B}^{-1}\left(\overline{\mathbf{x}_0^a}-\overline{\mathbf{x}_0^f}\right)+\frac{1}{2}\sum_t\left[H\left(\overline{\mathbf{x}_t^a}\right)-\mathbf{y}_t\right]^T\mathbf{R}_t^{-1}\left[H\left(\overline{\mathbf{x}_t^a}\right)-\mathbf{y}_t\right]$

Components formulation: $J=\frac{1}{2}\sum_{i_1,i_2}\left(\mathbf{B}^{-1}\right)_{i_1i_2}\left(\overline{x_{i_1,0}^a}-\overline{x_{i_1,0}^f}\right)\left(\overline{x_{i_2,0}^a}-\overline{x_{i_2,0}^f}\right)+\frac{1}{2}\sum_{k,t}\frac{1}{R_{k,t}}\left[H_k\left(\overline{\mathbf{x}_t^a}\right)-y_{k,t}\right]^2$

$$B_{i_1i_2}=\frac{1}{M-1}\sum_j \delta x_{i_1j,0}^f\,\delta x_{i_2j,0}^f$$

Approximation of B using ensemble

$$\overline{\mathbf{x}_t^a}=\overline{\mathbf{x}_t^f}+\sum_j \delta\mathbf{x}_{j,t}^f w_j$$

Weighted summation of ensemble perturbations

$\delta\mathbf{x}_{j,t}=\mathbf{x}_{j,t}^f-\overline{\mathbf{x}_t^f}$

is added to first guess

$$J=\frac{M-1}{2}\sum_j w_j^2+\frac{1}{2}\sum_{k,t}\frac{1}{R_{k,t}}\left[H_k\left(\overline{\mathbf{x}_t^a}\right)-y_{k,t}\right]^2$$

$$\frac{\partial J}{\partial w_j}=(M-1)w_j+\sum_{k,t}\frac{1}{R_{k,t}}\delta H_{kj,t}\left[H_k\left(\overline{\mathbf{x}_t^a}\right)-y_{k,t}\right]=0$$

Here, $\delta H_{kj,t}=H_k\left(\overline{\mathbf{x}_t^a}+\delta\mathbf{x}_{j,t}^f\right)-H_k\left(\overline{\mathbf{x}_t^a}\right)$

Cf: Zupanski et al. (2008)

$\Rightarrow$ **Solve for** w_j **and gain** $\overline{x_{i,t}^a}$

i: grid points(1-N)

k: obs. points(1-K)

t: time slots(1-T)

j: members(1-M)

EnVAR with observation localization

The number of ensemble members is usually too small to make analysis with large degree of freedom.

⇒ Localization required for increasing degree of freedom of w_j

Multiplicative inflation parameter

$$B_{i_1 i_2} = \frac{1}{M-1}\sum_j \delta x^f_{i_1 j,0}\delta x^f_{i_2 j,0} \Rightarrow B_{i_1 i_2} = \frac{\alpha^2}{M-1}\sum_j L_{i_1 i_2}\delta x^f_{i_1 j,0}\delta x^f_{i_2 j,0} \qquad L_{i_1 i_2} \equiv \sum_l L^{1/2}_{i_1 l} L^{1/2}_{i_2 l}$$

$$\overline{x^a_{i,t}} = \overline{x^f_{i,t}} + \sum_j \delta x^f_{ij,t} w_j \Rightarrow \overline{x^a_{i,t}} = \overline{x^f_{i,t}} + \sum_{l,j} L^{1/2}_{il}\delta x^f_{ij,t}\tilde{w}_{lj}$$

Localization factor (If grid i_1 is far from grid i_2, it is small or 0.)

Then, cost function:

$$\tilde{J} = \frac{M-1}{2\alpha^2}\sum_{l,j}\tilde{w}_{lj}^2 + \frac{1}{2}\sum_{k,t}\frac{1}{R_{k,t}}\left[H_k\left(\overline{\mathbf{x}^a_t}\right) - y_{k,t}\right]^2$$

Gradient of cost function:

$$\frac{\partial \tilde{J}}{\partial \tilde{w}_{lj}} = \frac{M-1}{\alpha^2}\tilde{w}_{lj} + \sum_{k,t}\frac{1}{R_{k,t}}\delta H_{lkj,t}\left[H_k\left(\overline{\mathbf{x}^a_t}\right) - y_{k,t}\right] = 0$$

How is this calculated?

i, l: grid points(1-N)

k: obs. points(1-K)

t: time slots(1-T)

j: members(1-M)

EnVAR with observation localization

$$\frac{\partial \tilde{J}}{\partial \tilde{w}_{lj}} = \frac{M-1}{\alpha^2}\tilde{w}_{lj} + \sum_{k,t}\frac{1}{R_{k,t}}\delta H_{lkj,t}\left[H_k\left(\overline{\mathbf{x}_t^a}\right) - y_{k,t}\right] = 0$$

Here,

$$\delta H_{lkj,t} = \frac{\partial H_k\left(\overline{\mathbf{x}_t^a}\right)}{\partial \tilde{w}_{lj}} = \sum_{i_1}\frac{\partial H_k\left(\overline{\mathbf{x}_t^a}\right)}{\partial \overline{x_{i_1,t}^a}}\frac{\partial \overline{x_{i_1,t}^a}}{\partial \tilde{w}_{lj}}$$

$$\approx \sum_{i_1}\frac{H_k\left(\overline{x_{1,t}^a},\ldots,\overline{x_{i_1,t}^a}+\delta x_{i_1 j,t}^f,\ldots,\overline{x_{N,t}^a}\right) - H_k\left(\overline{x_{1,t}^a},\ldots,\overline{x_{i_1,t}^a},\ldots,\overline{x_{N,t}^a}\right)}{\cancel{\delta x_{i_1 j,t}^f}} L_{i_1 l}^{1/2}\cancel{\delta x_{i_1 j,t}^f}$$

$$= \sum_{i_1} L_{i_1 l}^{1/2}\left[H_k\left(\overline{x_{1,t}^a},\ldots,\overline{x_{i_1,t}^a}+\delta x_{i_1 j,t}^f,\ldots,\overline{x_{N,t}^a}\right) - H_k\left(\overline{x_{1,t}^a},\ldots,\overline{x_{i_1,t}^a},\ldots,\overline{x_{N,t}^a}\right)\right]$$

$$\approx L_{kl}^{1/2}\left[H_k\left(\overline{x_{1,t}^a}+\delta x_{1j,t}^f,\ldots,\overline{x_{N,t}^a}+\delta x_{Nj,t}^f\right) - H_k\left(\overline{x_{1,t}^a},\ldots,\overline{x_{N,t}^a}\right)\right]$$

Using following equation

$$\overline{x_{i,t}^a} = \overline{x_{i,t}^f} + \sum_{l,j} L_{il}^{1/2}\delta x_{ij,t}^f \tilde{w}_{lj}$$

If i_1 and k are completely on the same point, it is same as $L_{i_1 l}^{1/2}$.
(Then, H_k depends only on the valuable on the grid i_1.)

i, l: grid points(1-N)

k: obs. points(1-K)

t: time slots(1-T)

j: members(1-M)

How to solve ∇J=0

$$\frac{\partial \tilde{J}}{\partial \tilde{w}_{lj}} = \frac{M-1}{\alpha^2}\tilde{w}_{lj} + \sum_{k,t}\frac{1}{R_{k,t}}L_{kl}^{1/2}\delta H_{kj,t}\left[H_k\left(\overline{\mathbf{x}_t^a}\right) - y_{k,t}\right] = 0$$

$$\overline{x_{i,t}^a} = \overline{x_{i,t}^f} + \sum_{l,j} L_{il}^{1/2}\delta x_{ij,t}^f \tilde{w}_{lj}$$

Not independent for analysis points

Using following equation

$$L_{i_1 i_2} \equiv \sum_l L_{i_1 l}^{1/2} L_{i_2 l}^{1/2}$$

$$w_{ij} \equiv \sum_l L_{il}^{1/2}\tilde{w}_{lj}$$

$$\frac{\partial J}{\partial w_{ij}} \equiv \sum_l L_{il}^{1/2}\frac{\partial \tilde{J}}{\partial \tilde{w}_{lj}} = \frac{M-1}{\alpha^2}w_{ij} + \sum_{k,t}\frac{1}{R_{k,t}}L_{ik}\delta H_{kj,t}\left[H_k\left(\overline{\mathbf{x}_t^a}\right) - y_{k,t}\right] = 0$$

Observation localization

$$\overline{x_{i,t}^a} = \overline{x_{i,t}^f} + \sum_j \delta x_{ij,t}^f w_{ij}$$

Independently calculated for every analysis points

Approximated new cost function is defined

$$J \cong \sum_i \left\{\frac{M-1}{2\alpha^2}\sum_j w_{ij}^2 + \frac{1}{2}\sum_{k,t}\frac{L_{ik}}{R_{k,t}}\left[H_k\left(\overline{\mathbf{x}_t^a}\right) - y_{k,t}\right]^2\right\}$$

i, l: grid points(1-N)

k: obs. points(1-K)

t: time slots(1-T)

j: members(1-M)

Summary of EnVAR formulation

i: analysis points
k: observation points
t: time slots
j: ensemble members

Analysis

$$\overline{x^a_{i,t}} = \overline{x^f_{i,t}} + \sum_j \delta x^f_{ij,t} w_{ij}$$

$$J = \frac{1}{2}\sum_i \left\{ \frac{M-1}{\alpha^2} \sum_j w_{ij}^2 + \sum_{k,t} \frac{L_{ik}}{R_{k,t}} \left[H_k\left(\overline{\mathbf{x}^a_t}\right) - y_{k,t} \right]^2 \right\}$$

Localization facter

Observation operator

Observation error variance

$$\frac{\partial J}{\partial w_{ij}} = \frac{M-1}{\alpha^2} w_{ij} + \sum_{k,t} \frac{L_{ik}}{R_{k,t}} \delta H_{kj,t} \left[H_k\left(\overline{\mathbf{x}^a_t}\right) - y_{k,t} \right] = 0$$

Minimization of globally defined J with observation localization

$$\delta H_{kj,t} = H_k\left(\overline{\mathbf{x}^a_t} + \delta \mathbf{x}^f_{j,t}\right) - H_k\left(\overline{\mathbf{x}^a_t}\right)$$

Analysis perturbation

$$\delta x^a_{ij,t} = \sum_{j_1} \delta x^f_{ij_1,t} T_{ijj_1}$$

$$T_{ijj_1} = \sqrt{M-1} \sum_{j_2} \lambda_{j_2}^{-1/2} U_{jj_2} U_{j_1 j_2}$$

Eigenvalue decomposition of Hessian matrix of J

$$\frac{\partial^2 J}{\partial w_{ij} \partial w_{ij_1}} = \frac{M-1}{\alpha^2} \delta_{jj_1} + \sum_{k,t} \frac{L_{ik}}{R_{k,t}} \delta H_{kj,t} \delta H_{kj_1,t} = \sum_{j_2} \lambda_{ij_2} U_{ijj_2} U_{ij_1 j_2}$$

Difference between LETKF and EnVAR

[1] H (Local or Global)

LETKF $H_{ik}\left(\overline{\mathbf{x}_t^a}\right) = H_k\left(\overline{x^f_{i_k,t}}\right) + \sum_j \delta H_{kj,t} w_{ij}$

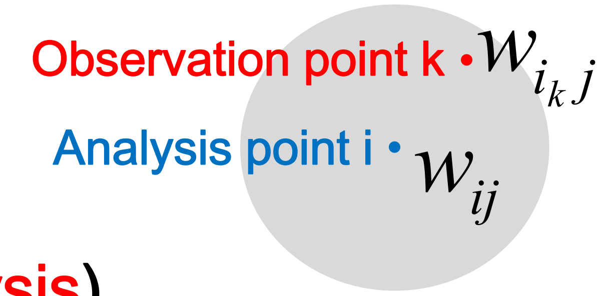
Calculated with w at analysis point
Independently for each analysis

EnVAR $H_k\left(\overline{\mathbf{x}_t^a}\right) = H_k\left(\overline{x^f_{i_k,t}} + \sum_j \delta x_{i_k j,t} w_{i_k j}\right)$

Calculated with w at observation point
for all analysis

EnVAR calculates H directly

If H is linear and analysis point is same as observation point, EnVAR=LETKF



[2] Gradient of H (around First guess or Analysis)

LETKF $\delta H_{kj,t} = H_k\left(\overline{\mathbf{x}_t^f} + \delta \mathbf{x}^f_{j,t}\right) - \overline{H_k\left(\mathbf{x}_t^f\right)}$

EnVAR $\delta H_{kj,t} = H_k\left(\overline{\mathbf{x}_t^a} + \delta \mathbf{x}^f_{j,t}\right) - H_k\left(\overline{\mathbf{x}_t^a}\right)$

δH in EnVAR is around analysis

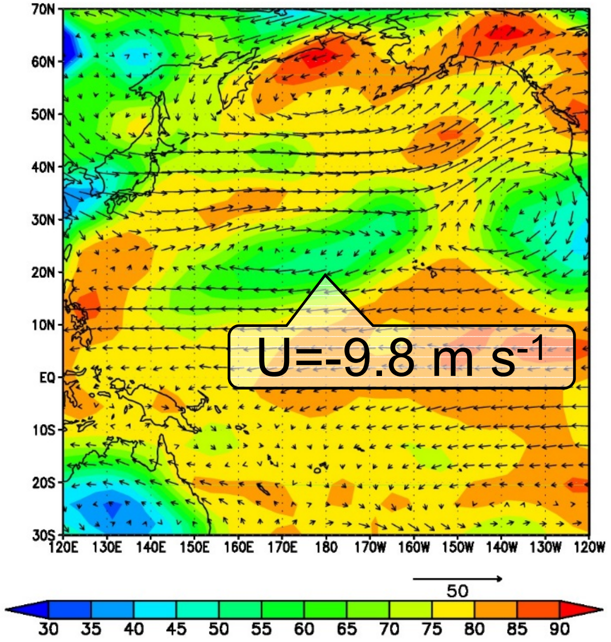
If H is linear, EnVAR=LETKF

i: analysis points
k: observation points
t: time slots
j: ensemble members

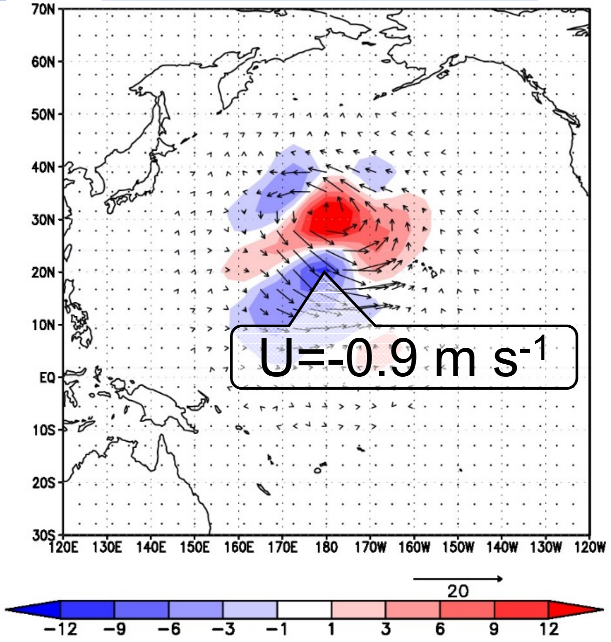
Comparison to LETKF (Linear H case)

Number of members: **20**
Localization radius: σ_H=1000(km), σ_V=0.1(sigma)
Observations: U=5m/s @20N,180E, σ=0.835

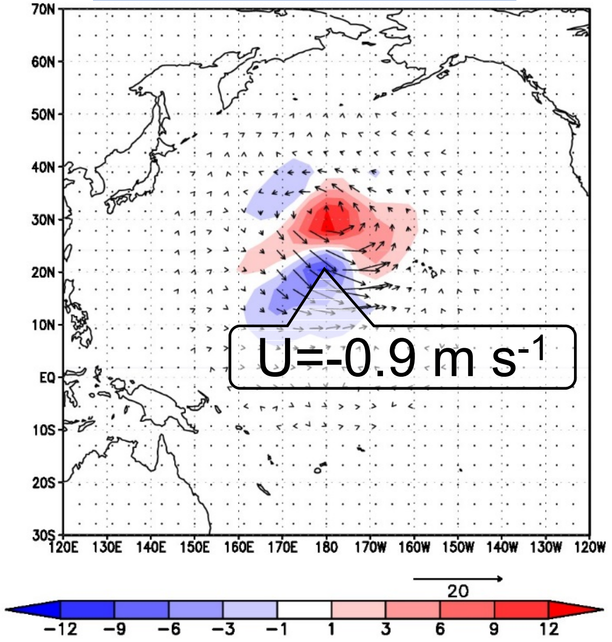
First guess of relative humidity and horizontal wind



Analysis-First guess (LETKF)



Analysis-First guess (EnVAR)



Increment of EnVAR is same as LETKF at observation point but smaller far from observation point ("severer" localization)

Why is EnVAR localization "severer" than LETKF?

Observation localization of this EnVAR is derived from background localization, but that of LETKF is not.

e.g., In two analysis points and one observation, $\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}\left[\mathbf{y} - H\left(\overline{\mathbf{x}^f}\right)\right]$, $H\left(\overline{\mathbf{x}^f}\right) = x_1$, and $\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} \dfrac{B_{11}}{B_{11}+R_1} \\ \dfrac{B_{12}}{B_{11}+R_1} \end{bmatrix}$

Background localization

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} \dfrac{B_{11}}{B_{11}+R_1} \\ \dfrac{LB_{12}}{B_{11}+R_1} \end{bmatrix}$$

Observation localization of LETKF

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} \dfrac{B_{11}}{B_{11}+R_1} \\ \dfrac{B_{12}}{B_{11}+R_1/L} \end{bmatrix} = \begin{bmatrix} \dfrac{B_{11}}{B_{11}+R_1} \\ \dfrac{LB_{12}}{LB_{11}+R_1} \end{bmatrix}$$

$$L < 1$$

Then, $K_2^{B-localization} < K_2^{R-localization}$

(Greybush et al. 2011)

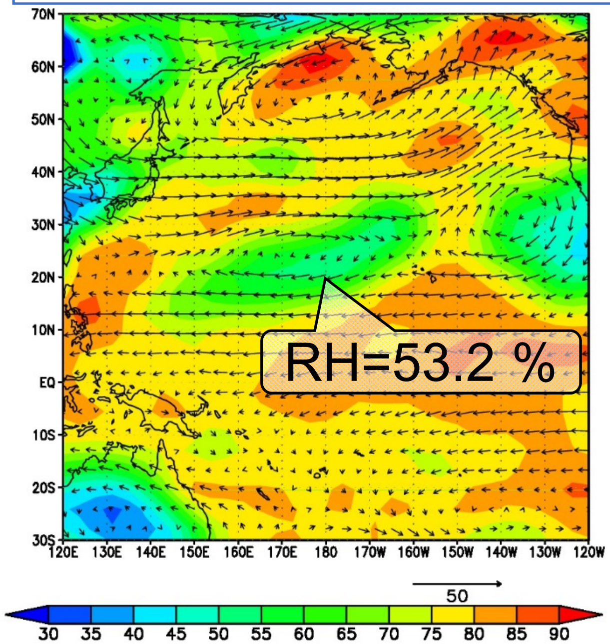
Therefore, background localization is "severer"

Single-observation assimilation

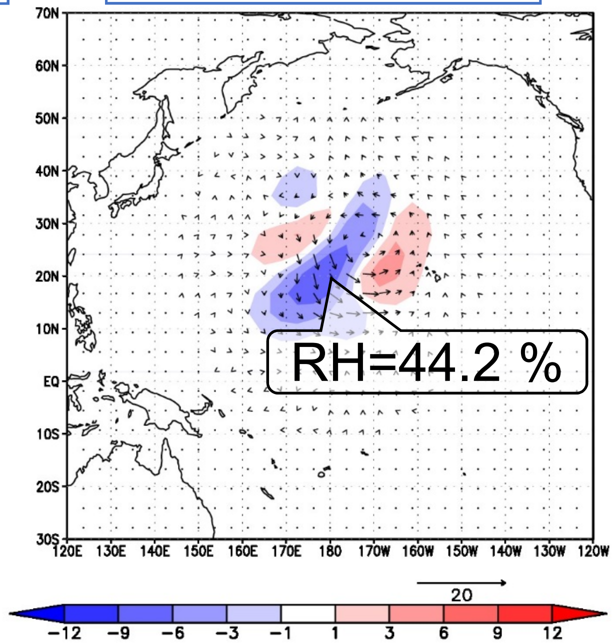
Comparison to LETKF (Non-linear H case)

Number of members: **20**
Localization radius: **σ_H=1000(km), σ_V=0.1(sigma)**
Observations: RH=30% **@20N,180E, σ=0.835**

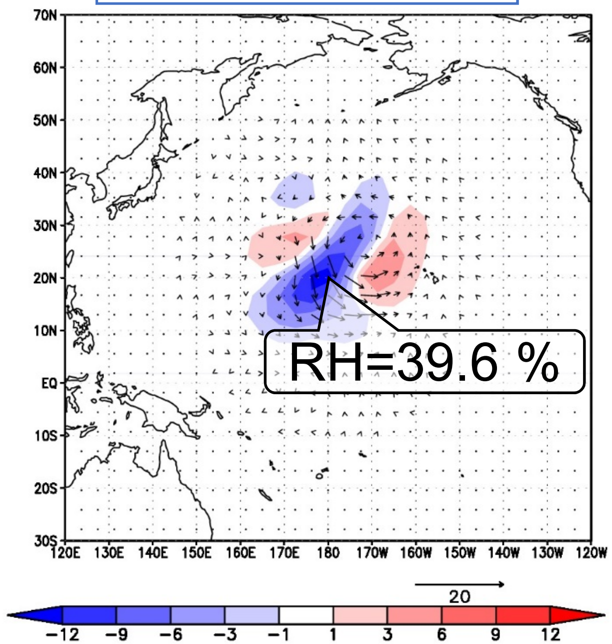
First guess of relative humidity and horizontal wind



Analysis-First guess (LETKF)



Analysis-First guess (EnVAR)

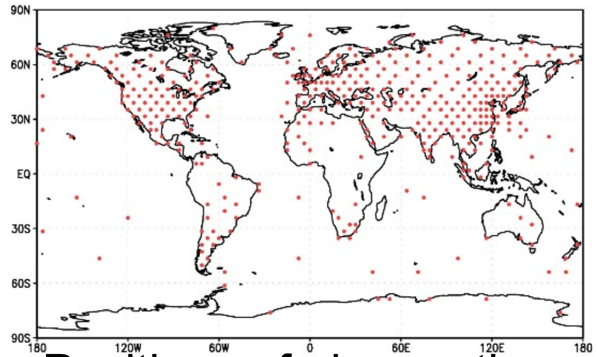


EnVAR analysis is closer to observation than LETKF

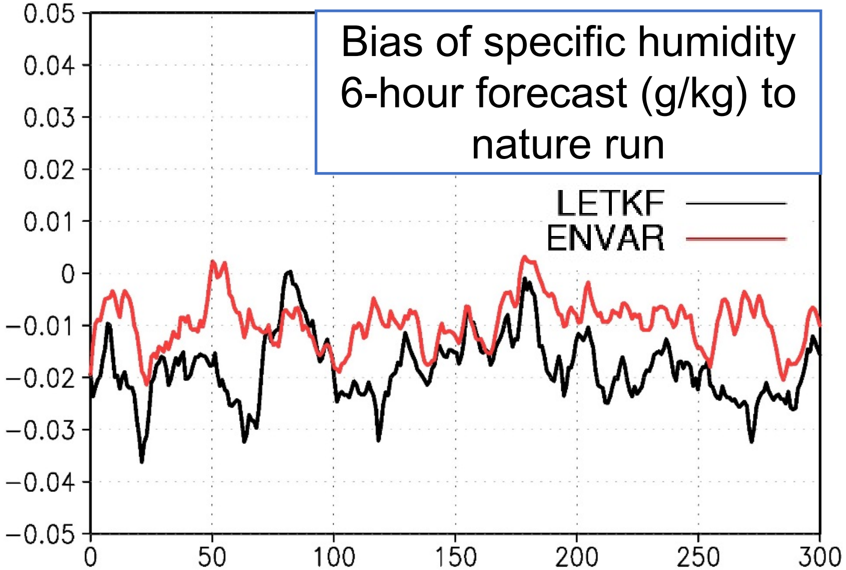
OSSE

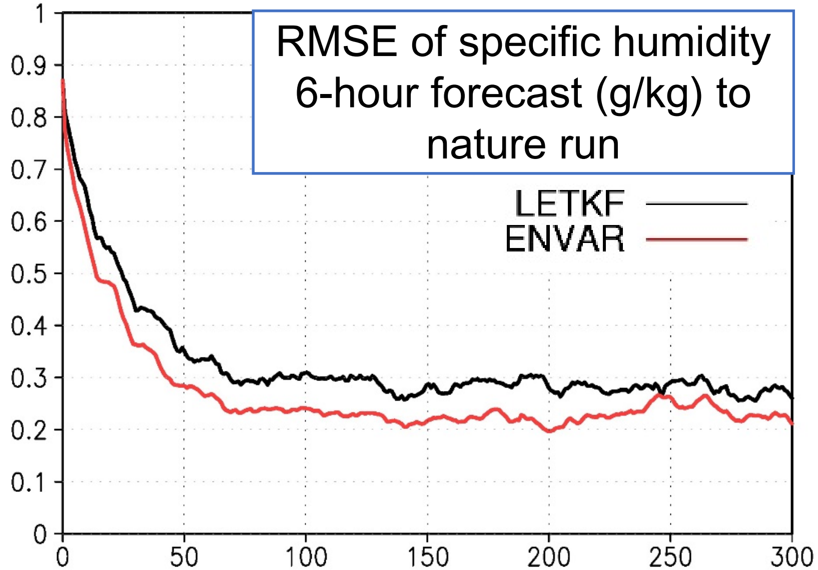
Observation system simulation experiments with SPEEDY model

Number of members: **20**
Assimilation window: **6 hours**
Localization radius: **σ_H=1000(km), σ_V=0.1(sigma)**
Observations: **U, V, T, RH, Ps**
Inflation: **Multiplicative (=1.1)**
Analysis time: **center of the window**
Observation time: **t=1h, 3h, 5h**



Positions of observations





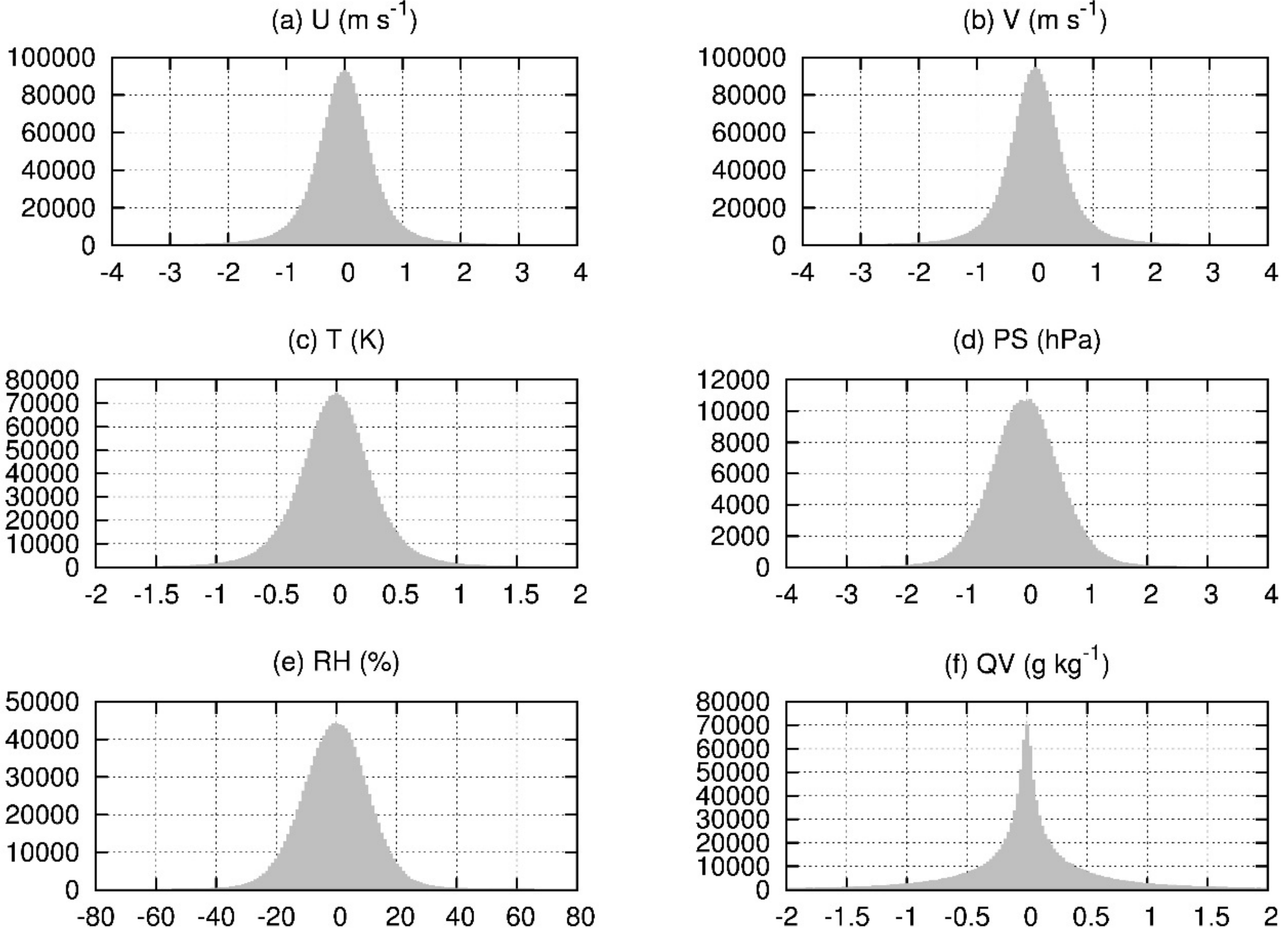
Number of forecast-analysis cycles (every 6 hour)

EnVAR is better caused by difference of how to calculate H and δH

OSSE

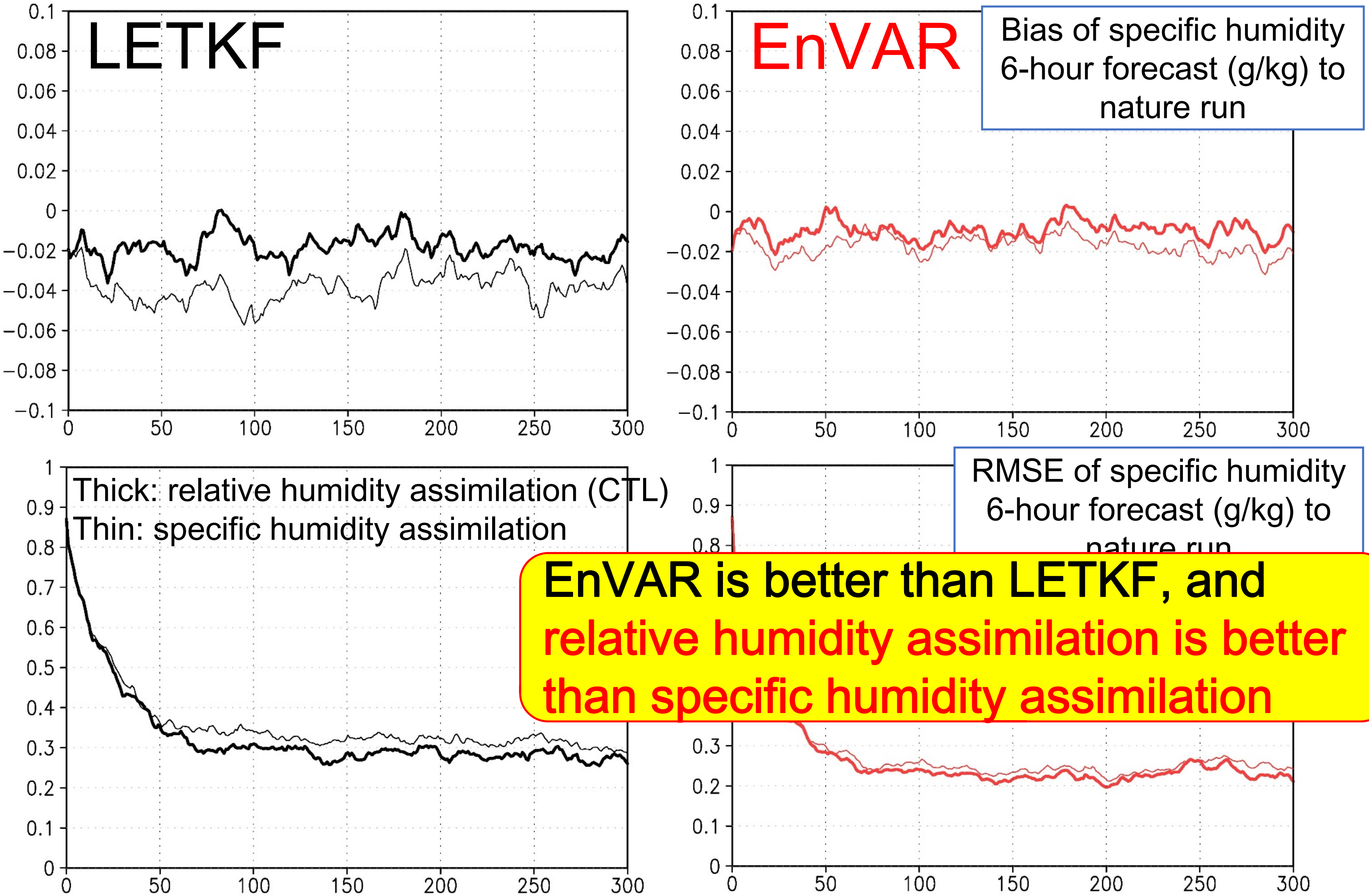
O-F Histgram

Observation-Forecast (O-F) histogram in all EnVAR analysis



Specific humidity assimilation:
Linear H but non-Gaussian probability distribution

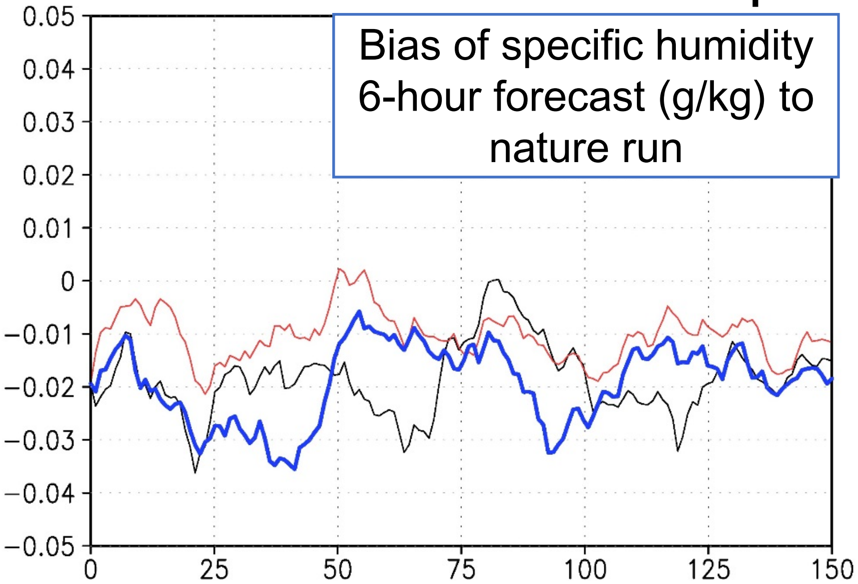
Specific humidity assimilation

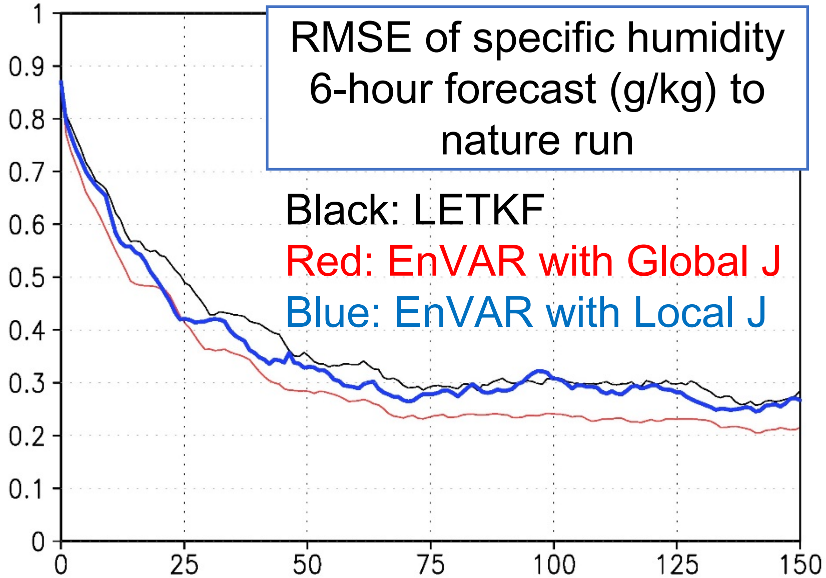


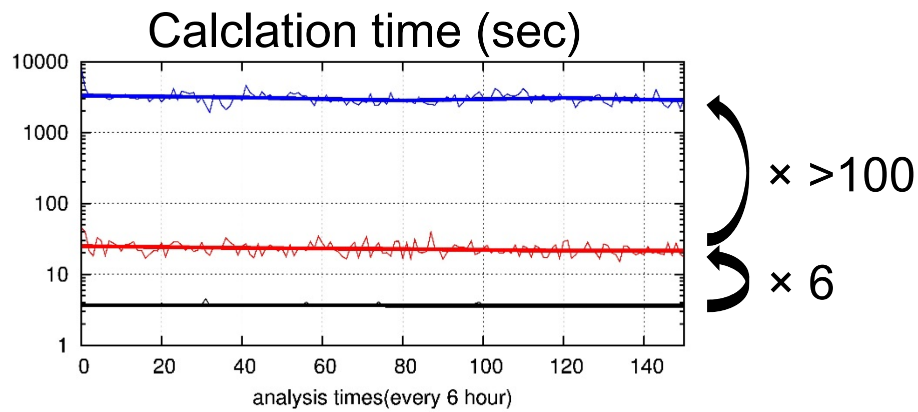
EnVAR is better than LETKF, and relative humidity assimilation is better than specific humidity assimilation

EnVAR with locally defined cost function

Local J is minimized implicitly







Globally defined H

$$H_k\left(\overline{\mathbf{x}_t^a}\right) = H_k\left(\overline{x_{i_k,t}^f} + \sum_j \delta x_{i_k j,t} w_{i_k j}\right)$$

calculated with w at observation point

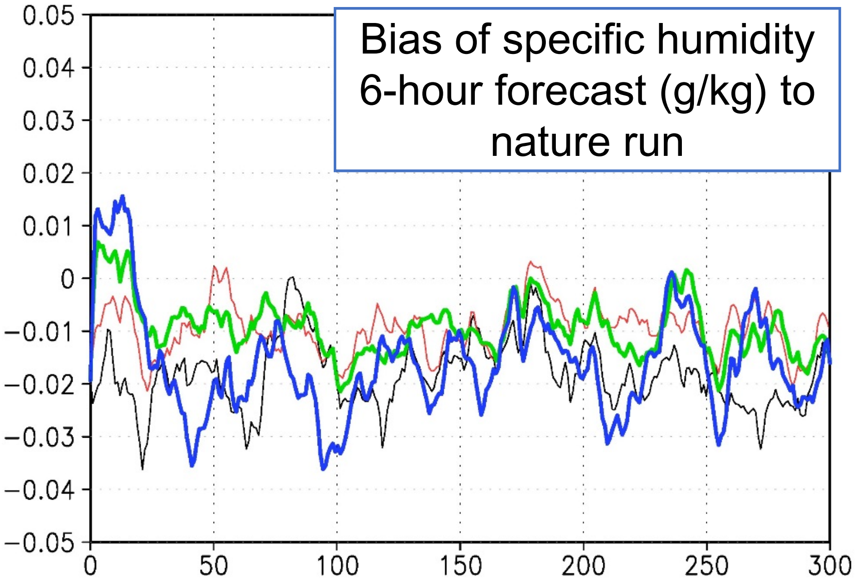
Locally defined H

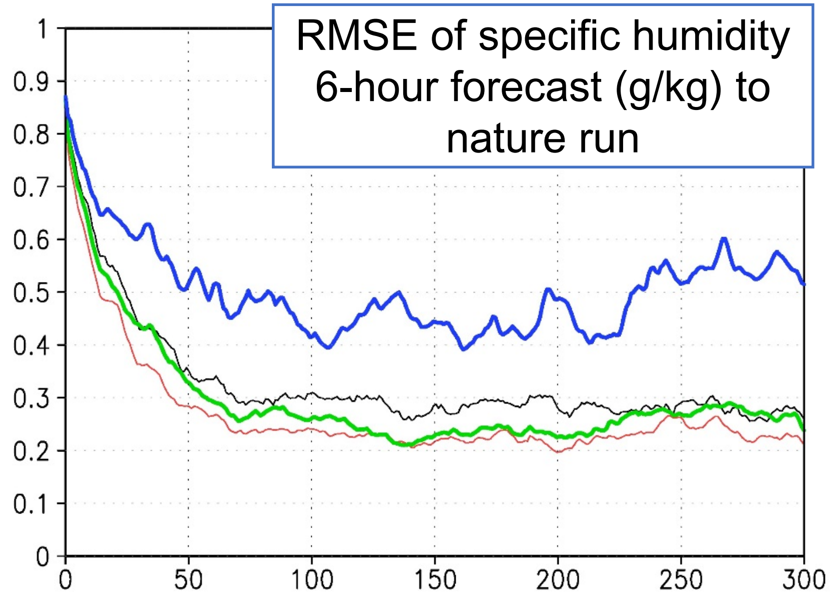
$$H_{ik}\left(\overline{\mathbf{x}_t^a}\right) = H_k\left(\overline{x_{i_k,t}^f} + \sum_j \delta x_{i_k j,t} w_{ij}\right)$$

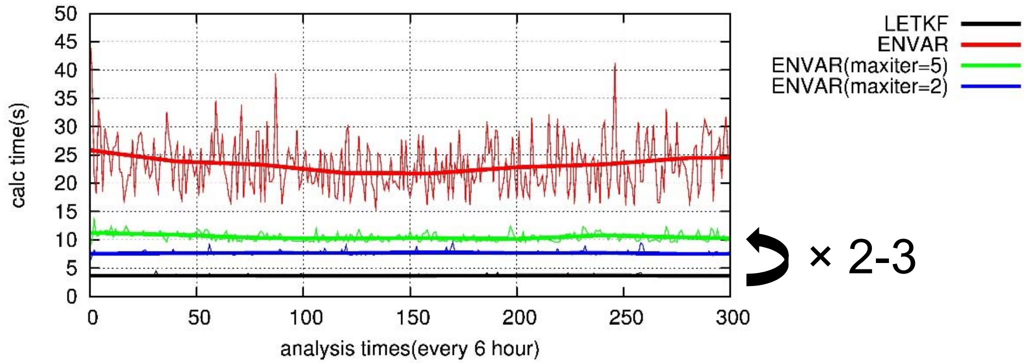
calculated with w at analysis point

Similar to LETKF
→ Global J has a good impact

EnVAR with the specific number of iteration







Black: LETKF

Red: EnVAR (CTL)

Green: EnVAR (stop after 5 iterations)

Blue: EnVAR (stop after 2 iterations)

EnVAR with 5 iterations is better than LETKF
(calculation time is 2-3 times as long as LETKF)

Summary of OSSEs

- We developed EnVAR with observation localization and compared it to LETKF
  - ◎ EnVAR analysis was closer to true value than LETKF because globally defined cost function is minimized
  - × Several times longer calculation time than LETKF

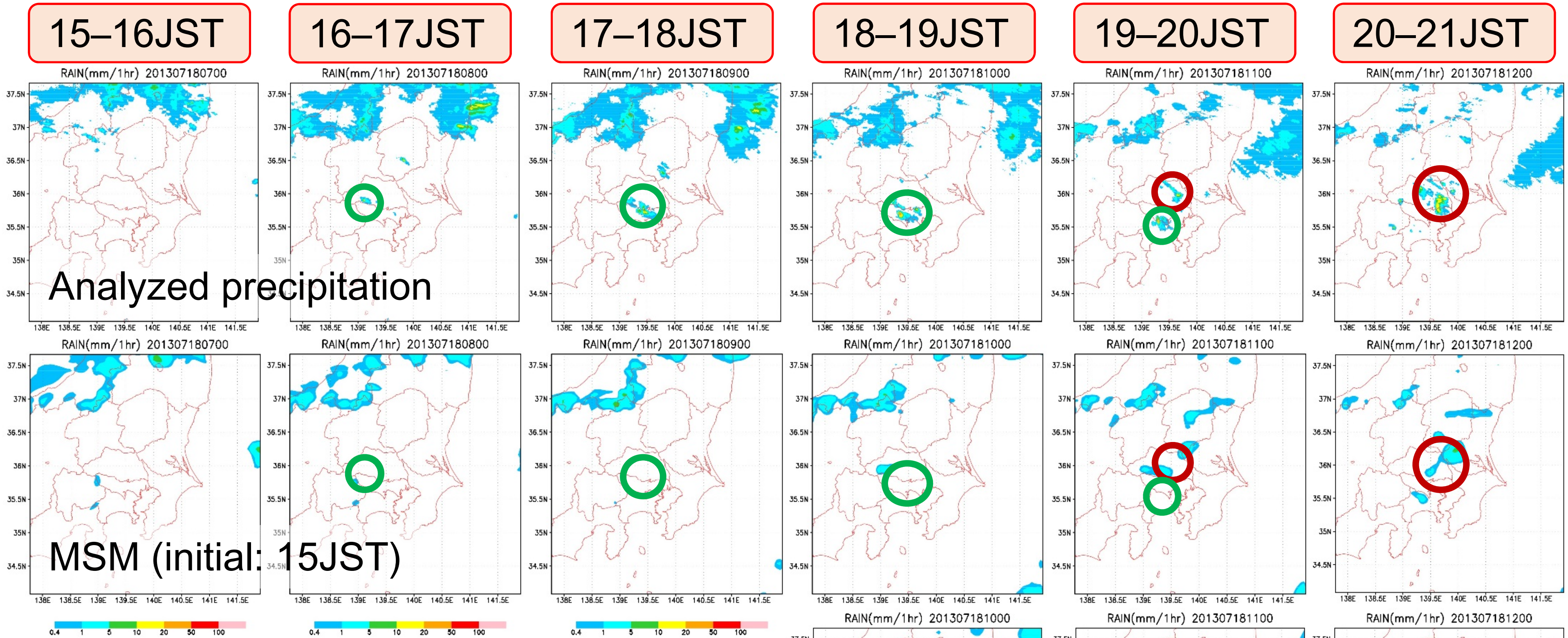
- Non-linear observation operator is strictly treated in EnVAR (Gaussian approximation should not be used)

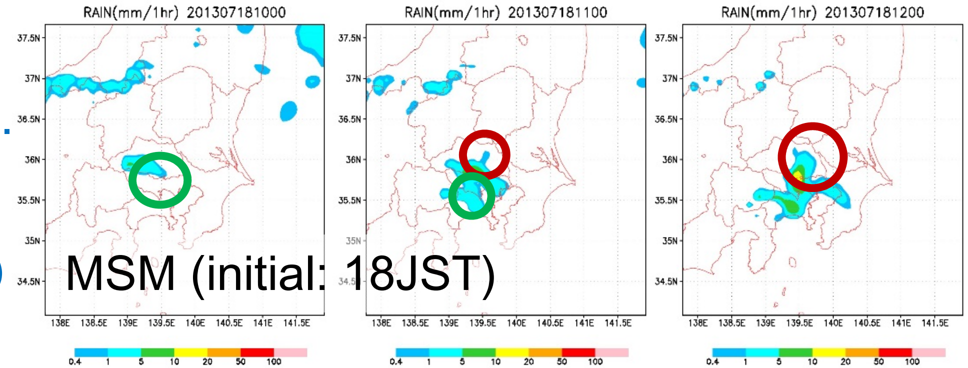
- Observation localization of this EnVAR is same as background localization ("severer" than localization of LETKF)

Is EnVAR also better than LETKF in real obs. data assimilation?

Real data assimilation

Local Rainfall on 18 July 2013





- Two precipitation systems were generated.
- Accurate forecast is difficult.
(even though the initial condition included rainfall)

Dense observations are expected to improve forecasts

Real data assimilation

Assimilated Dense Observations

Observation	Elements	Frequency	
Surface (JMA Surface observation and AMeDAS)	U, V, T	every 10 minutes	
GNSS	PWV	every 10 minutes	
Radar	Radial wind	every 10 minutes	Kashiwa, Haneda, Narita
Radiosonde	U, V, T, RH	every 3 hours	Tsukuba, Urawa, Yokosuka, Ryofu Maru

<u>Setting</u>

Horizontal localization: 20m

Vertical localization: 0.1 InP

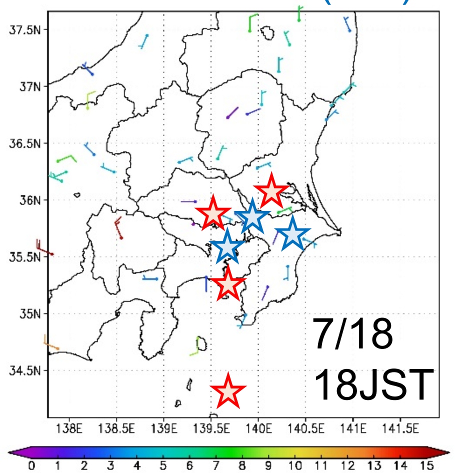
(PWV is not localized vertically)

Multiplicative inflation parameter: 1.2

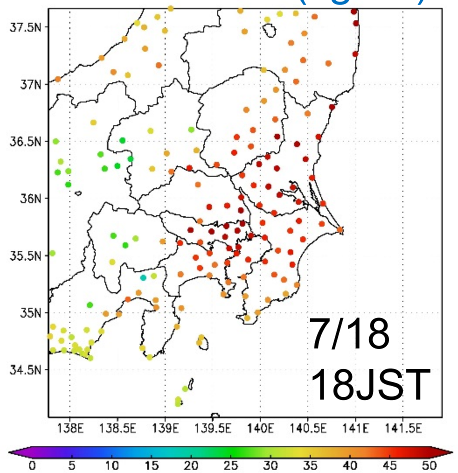
Observation error:

- U, V: 1 m/s
- T: 1 K
- RH: 10%
- PWV: 5 kg/m^2
- Radial wind: 3 m/s

Surface wind (m/s)



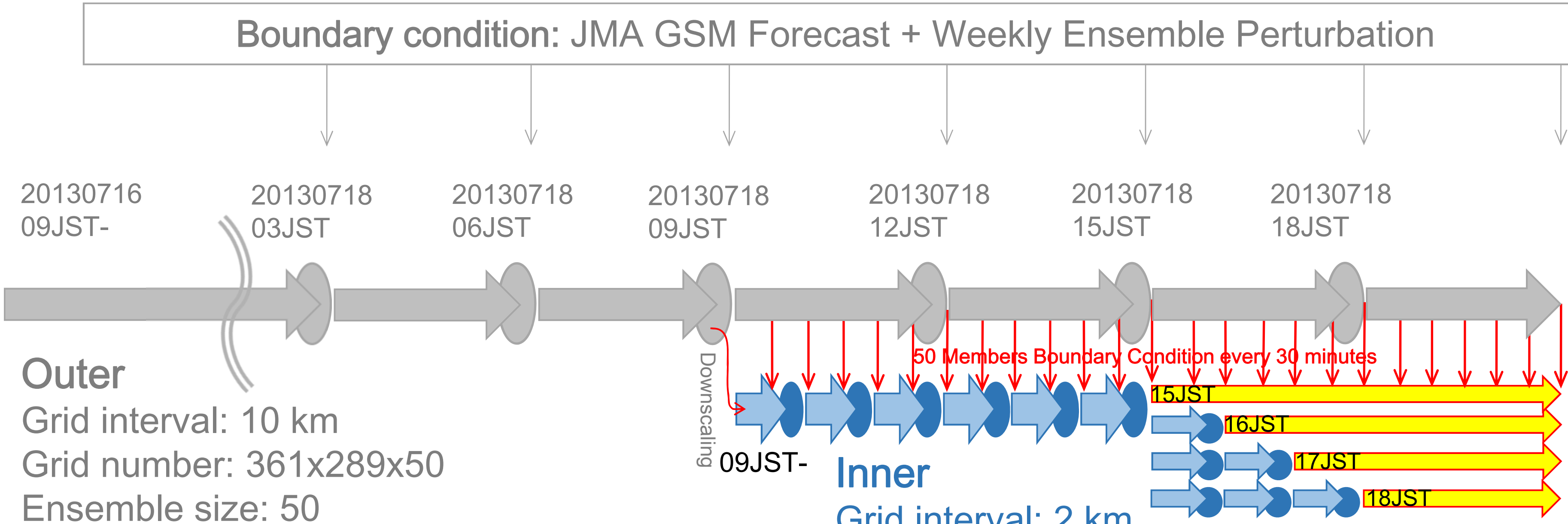
GNSS PWV (kg/m^2)



☆:Radar

☆:Sonde

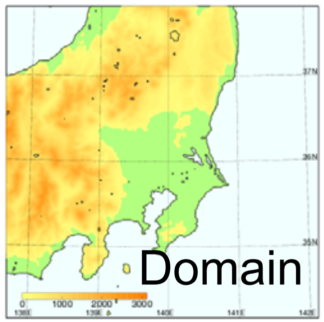
Flow of Assimilation Experiments



Outer
Grid interval: 10 km
Grid number: 361x289x50
Ensemble size: 50
Analysis window: 3 hour
(Operational Observations used in JMA Meso-DA every 30 minutes)

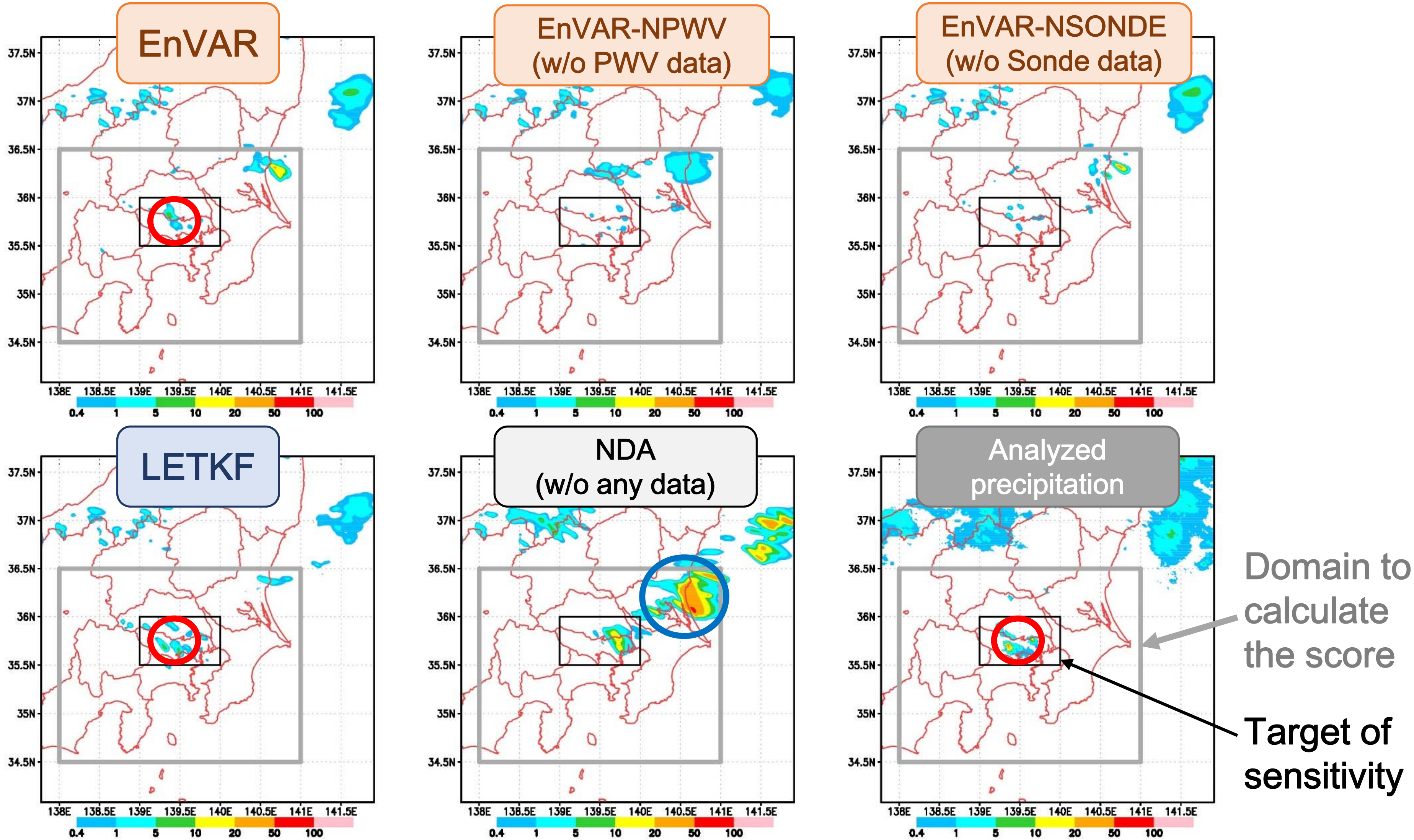
Inner
Grid interval: 2 km
Grid number: 200x200x60
Ensemble size: 50
Analysis window: 1 hour

: Ensemble Forecasts
: Analysis (LETKF or EnVAR)



Target:
Local rain near Tokyo in 20130718 17-21JST

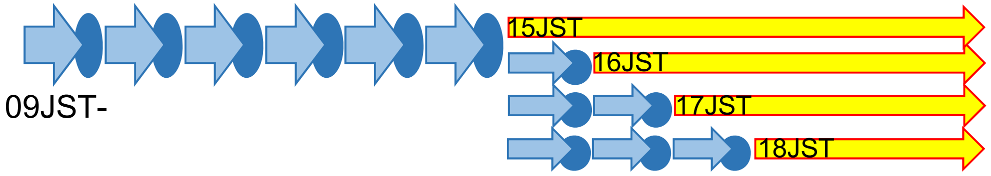
Comparison of 1-h Rainfall in 18-19 JST



Good impacts of PWV and Sonde data assimilation

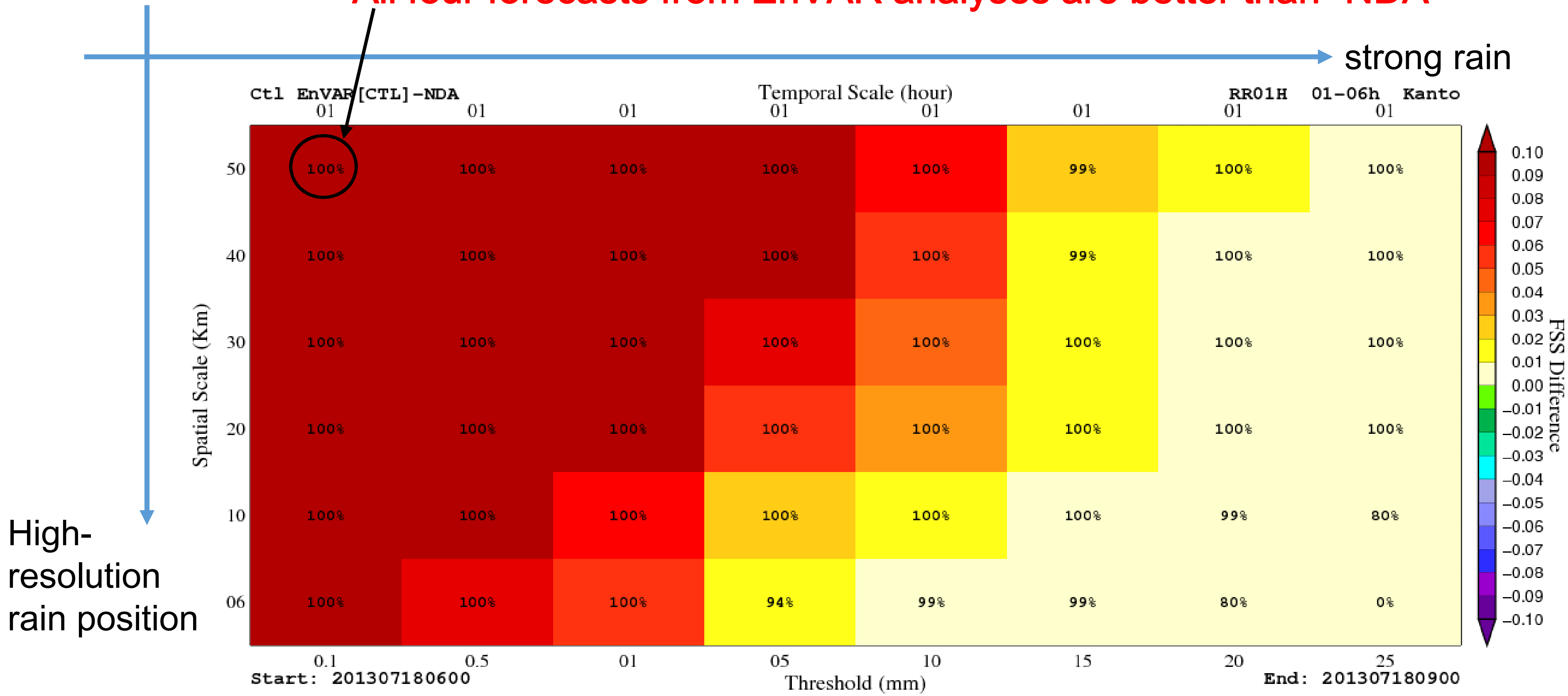
Are Fractions Skill Scores improved?

$$FSS = 1 - \frac{\left[\sum_i (O_i - F_i)\right]^2}{\left[\sum_i O_i\right]^2 + \left[\sum_i F_i\right]^2}$$



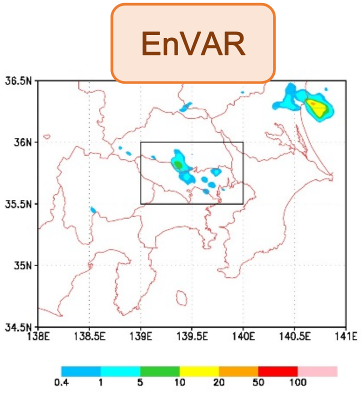
O_i : number density of <u>observed</u> rainfall in i-th fraction
F_i : number density of <u>forecast</u> rainfall in i-th fraction

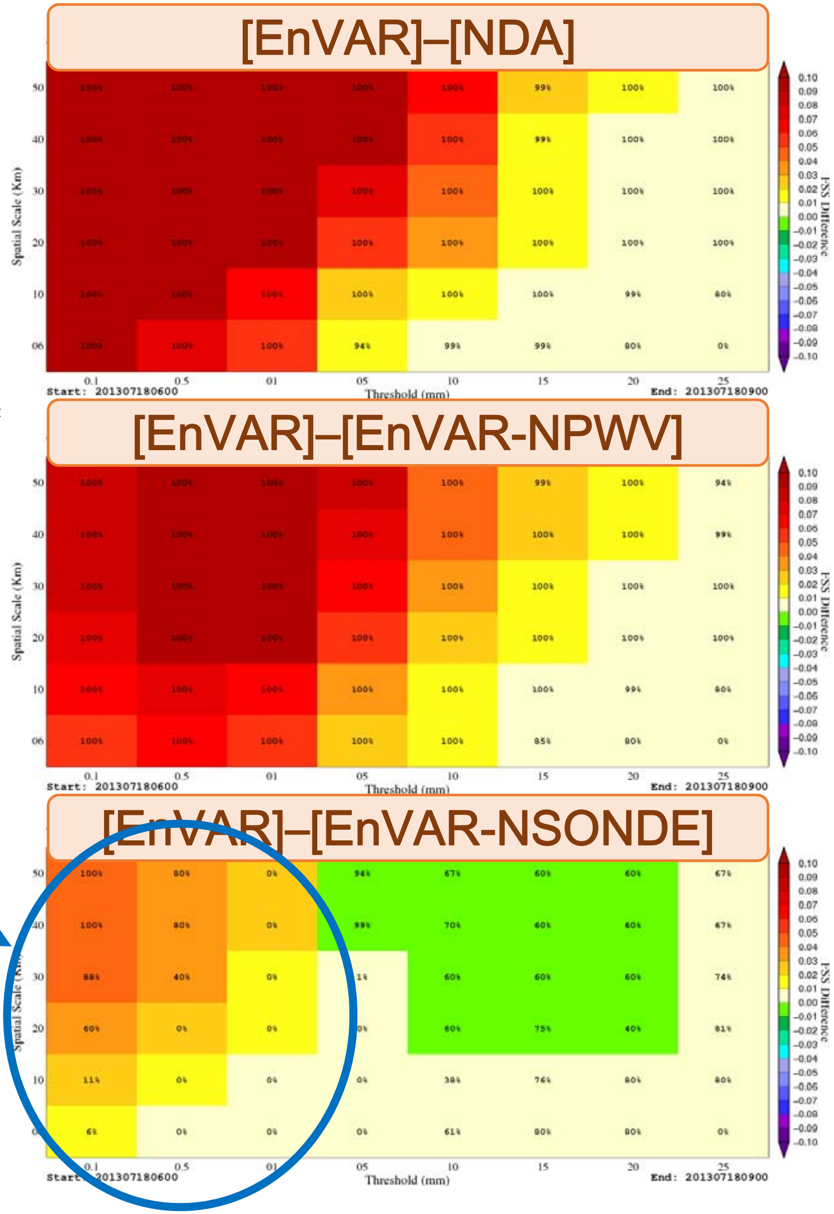
All four forecasts from EnVAR analyses are better than "NDA"



Impact of Dense Observations

Rainfall in 18-19 JST





- PWV data greatly improved rainfall forecasts.

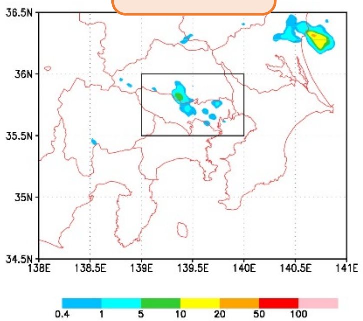
- Radiosonde data also improved weak rain forecasts.

Both PWV and radiosonde data could improve rainfall forecasts

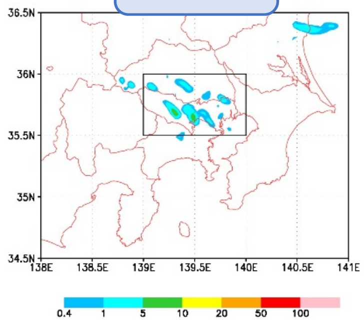
EnVAR v.s. LETKF

Rainfall in 18-19 JST

EnVAR

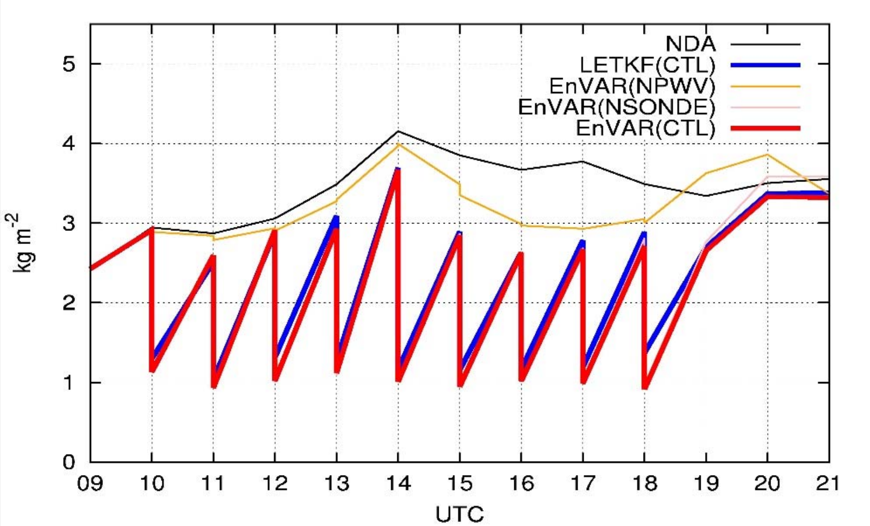


LETKF

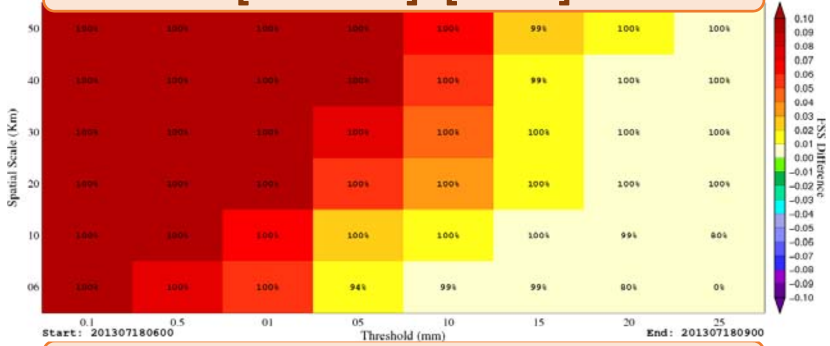


- Difference between EnVAR and LETKF is small

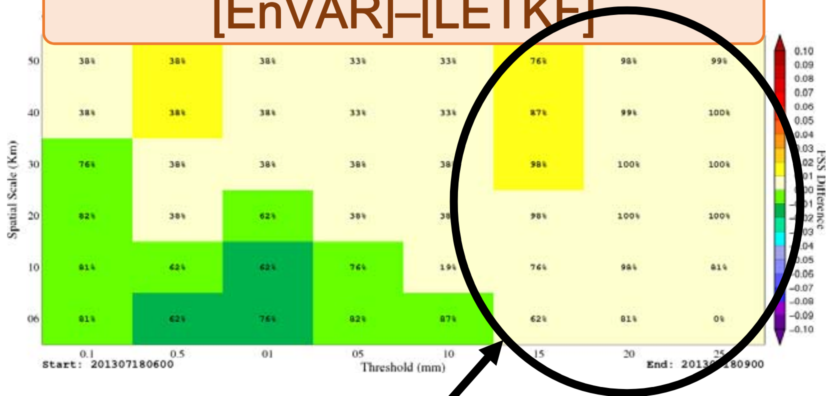
Time series of RMS of (O–A) and (O–F) of PWV in the forecast-analysis cycles



[EnVAR]–[NDA]



[EnVAR]–[LETKF]



In EnVAR, strong rainfall (> 15 mm/hr) forecasts are slightly better than that of LETKF

Correlation between Rainfall and Initial States

Correlation between J and x_n

$$\mathrm{CORR}(i,j)=\frac{\sum_m \left(J_m-\overline{J}\right)\left(x_m(i,j)-\overline{x_m(i,j)}\right)}{\sqrt{\sum_m \left(J_m-\overline{J}\right)^2}\sqrt{\sum_m \left(x_m(i,j)-\overline{x_m(i,j)}\right)^2}}$$

i, j : grid number, **m**: ensemble member

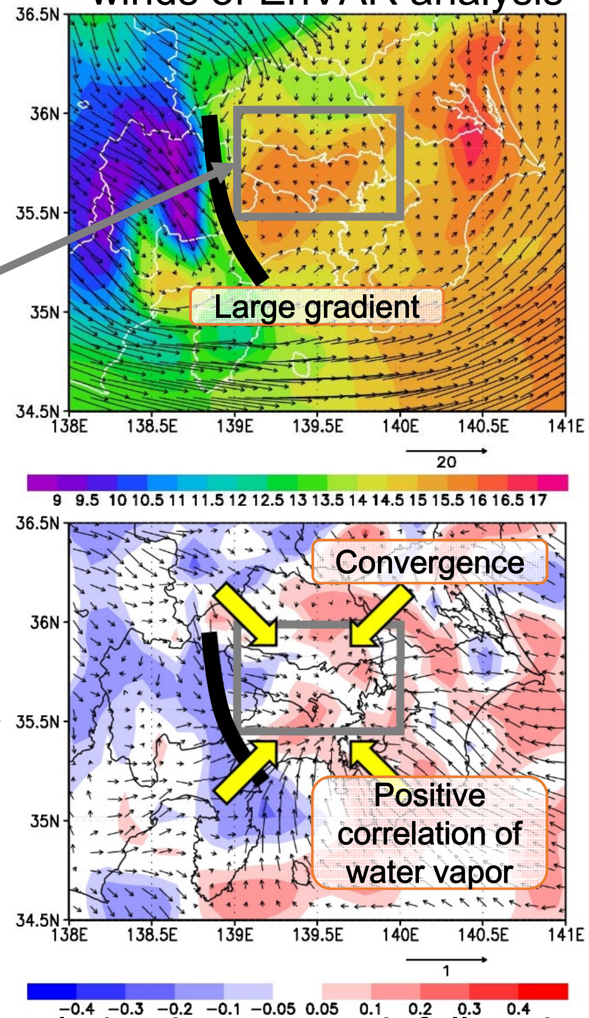
J_m : 1-h rainfall (18–19JST) averaged in this area

$x_m(i,j)$: variables in 0–1 km height at 18JST

If winds point to the direction of vectors in this figure, rainfall becomes stronger

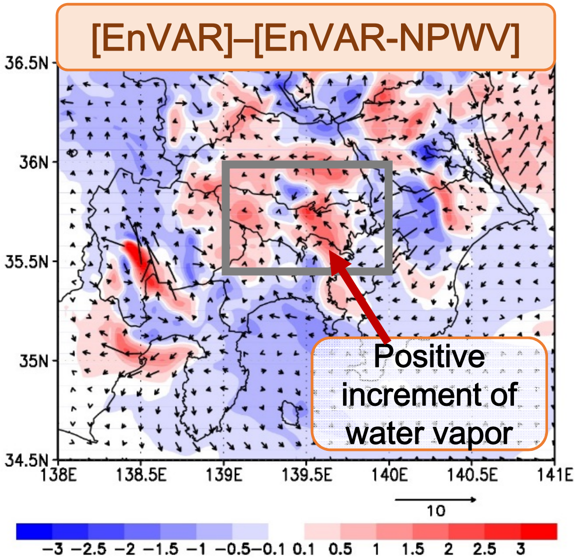
Low-level convergence is correlated to rainfall intensity

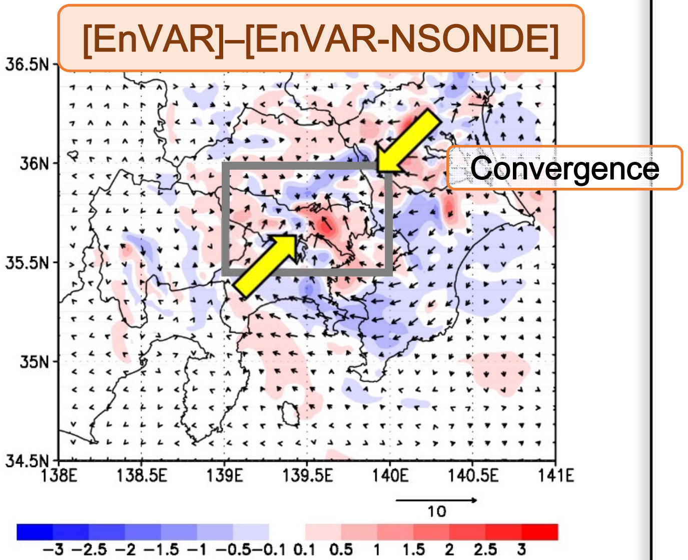
0–1 km water vapor and winds of EnVAR analysis

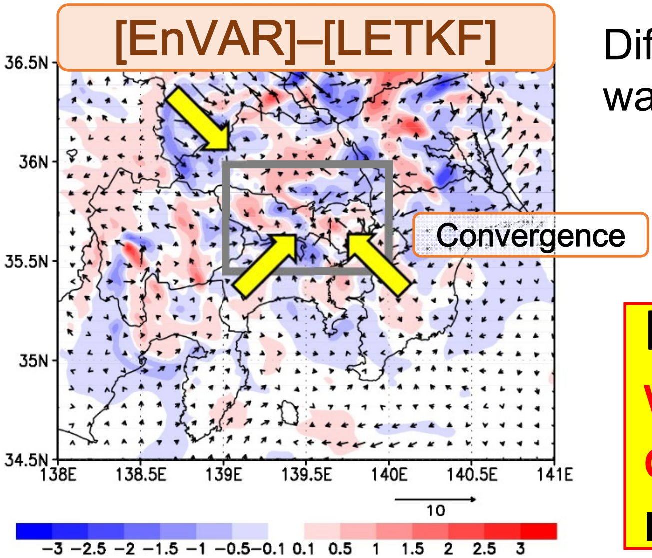


Correlation between rainfall and 0–1km water vapor and winds calculated by 51-member EnVAR

Difference of Low-level variables



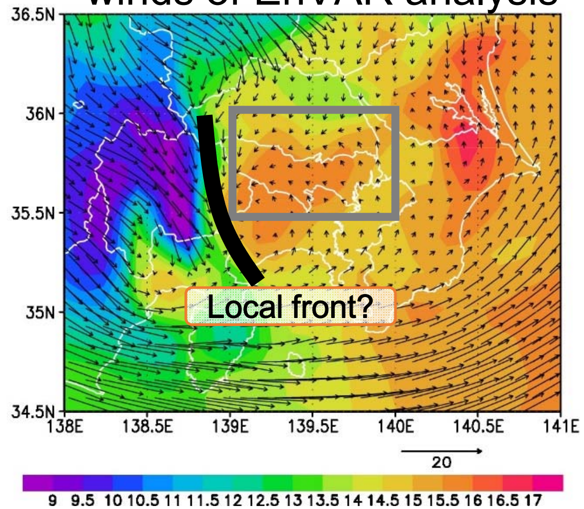


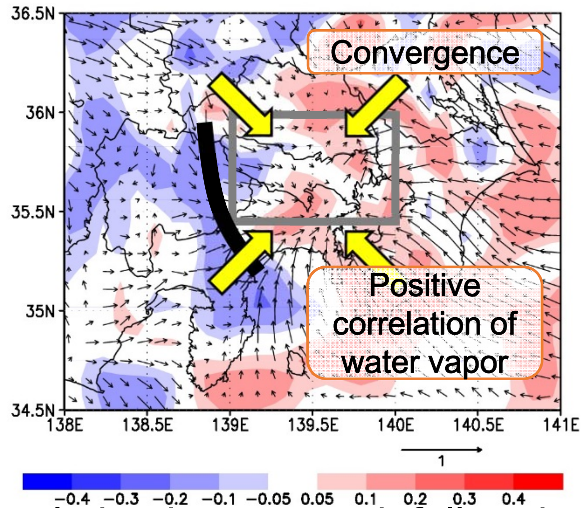


Difference of 0–1km water vapor and winds

Increment of low-level water vapor and convergence makes rainfall stronger

0–1 km water vapor and winds of EnVAR analysis





Correlation between rainfall and 0–1km water vapor and winds calculated by 51-member EnVAR

Summary of Real Data Assimilation

We assimilated dense obs. for the local rainfall near Tokyo

- Impact of dense PWV and Radiosonde obs.
  - PWV improved rainfall forecast through correcting low-level water vapor
  - Sonde obs. improved rainfall forecast through correcting low-level winds
- Comparison between LETKF and EnVAR
  - EnVAR can make the analysis which is closer to obs. than LETKF
  - Improvement of rainfall forecast by using EnVAR is small
- Correlation to rainfall based on ensemble forecasts
  - Low-level water vapor and convergence made local rainfall stronger

Are these impacts general? Verification in longer period requires

Summary (EnVAR v.s. LETKF)

- Formulation
  - Non-linear observation operator is more strictly treated in EnVAR
  - Observation localization of EnVAR is same as background localization ("severer" than localization of LETKF)

- OSSEs
  - Analysis of this EnVAR alization are more accurate than LETKF because globally defined cost function is minimized

- Real observation data assimilation
  - PWV assimilation in both EnVAR and LETKF improved rainfall forecast
  - EnVAR analysis was closer to obs. than LETKF
  - Improvement of rainfall forecast by using EnVAR was small in this case

Our research was supported in part by "Strategic Program for Innovative Research (SPIRE), Field 3" (proposal number: hp140220 and hp150214) and "Tokyo Metropolitan Area Convection Study for Extreme Weather Resilient Cities (TOMACS)". SPEEDY-LETKF (https://code.google.com/p/miyoshi/) and the source code developed by Numerical Prediction Division in JMA are used in this study. GNSS data were provided from the 2nd Laboratory, Meteorological Satellite and Observation System Research Department in MRI. Radiosonde observations were conducted as a part of TOMACS program. The other observation data were from JMA.