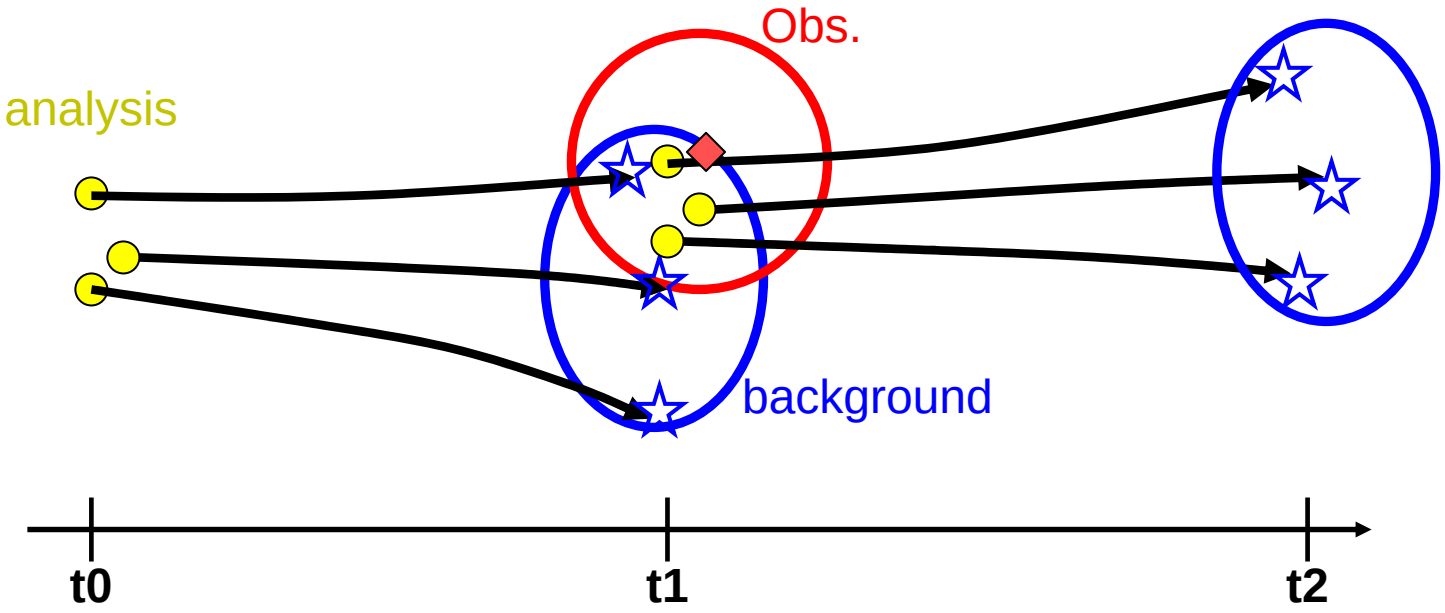
Applications of the LETKF to adaptive observations, analysis sensitivity, observation impact and assimilation of moisture

Junjie Liu and Eugenia Kalnay

DEPARTMENT OF
ATMOSPHERIC &
OCEANIC SCIENCE

Dec. 4th, 2007

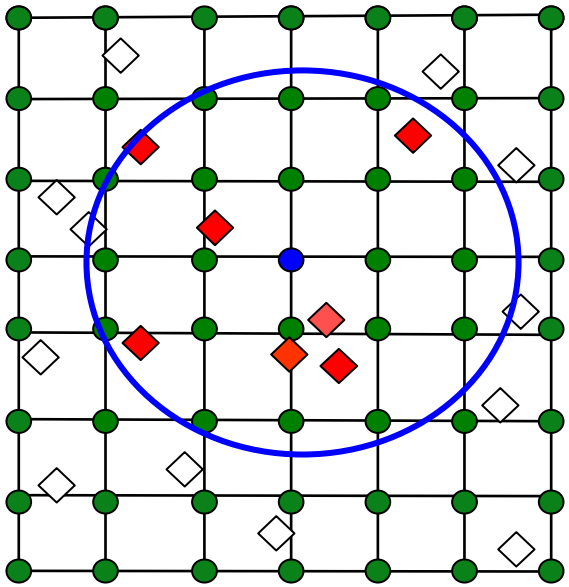
Local Ensemble Transform Kalman Filter
(LETKF, Ott et al., 2004, Hunt et al., 2007)



LETKF, like any other EnKF, provides background and analysis uncertainty estimation in every analysis cycle.

Local Ensemble Transform Kalman Filter
(LETKF, Ott et al., 2004, Hunt et al., 2007)

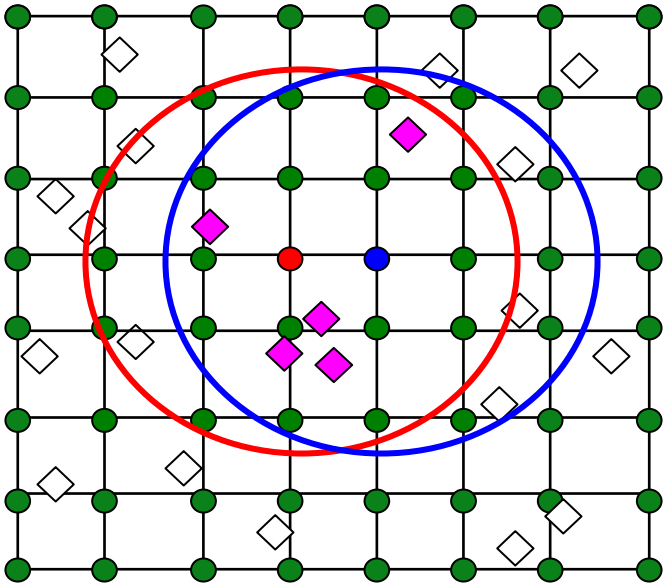
Schematic of 2-dimension local patch



✓ LETKF solves the analysis states in a local volume centered each grid point

Local Ensemble Transform Kalman Filter
(LETKF, Ott et al., 2004, Hunt et al., 2007)

Schematic of 2-dimension local patch



- Different local volumes have a great overlap.
- Each observation is used more than once in the data assimilation.
- The analysis in each grid point is highly parallel.

Outline

- New applications of the LETKF
  - Adaptive observations
  - Analysis sensitivity and information content
  - Observation impact
  - Assimilation of moisture
- Summary

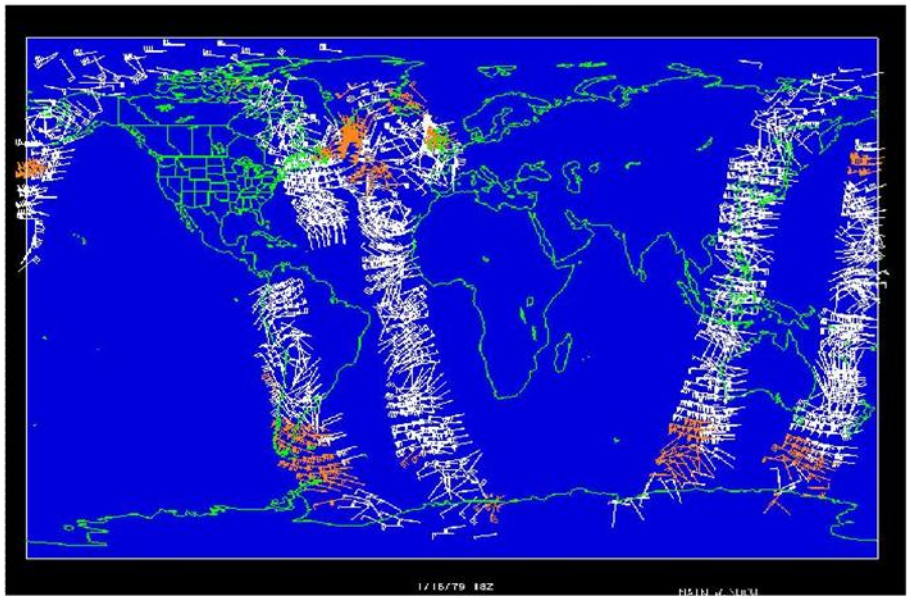
Outline

- New applications of the LETKF
  - Adaptive observations (Liu and Kalnay, GRL, 2007)
  - Analysis sensitivity and information content
  - Observation impact
  - Assimilation of moisture
- Summary

Motivation of adaptive observations

EXAMPLE TARGETED LOCATIONS FOR DWL OSSE

(White symbols: full lidar coverage; Red symbols: targeted coverage)



White: DWL orbits

Brown: adaptive DWL

Hardesty, 2006

• U. S. Doppler Wind Lidar (DWL) will be operated in targeted mode, with the goal of "10% adaptive observations to get 90% improvement"

MY GOAL: sample simulated DWL observations based on LETKF estimated uncertainty and compare it to other sampling strategies.

Experimental Design

- Simplified PrimitivE Equation DYnamics model (SPEEDY) (Molteni, 2003, adapted by Miyoshi, 2005)
  - A global model with fast computation speed.
  - 96 grid points zonally, 48 grid points meridionally, and 7 vertical levels

- Data assimilation schemes
  - Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007)
  - 3D-Var (Miyoshi, 2005)

- Simulated observations (perfect model assumption)
  - "Truth" (a long time integration) plus random perturbations.
  - Observe both zonal and meridional wind at adaptive observation points.

Sampling strategies

- Ensemble spread strategy (from the LETKF)

  → Locations with large ensemble wind spread at 500hPa.
- Random locations
- Uniform distribution
- Climatology ensemble spread

  → Locations with large climatological average ensemble wind spread from rawinsonde assimilation.

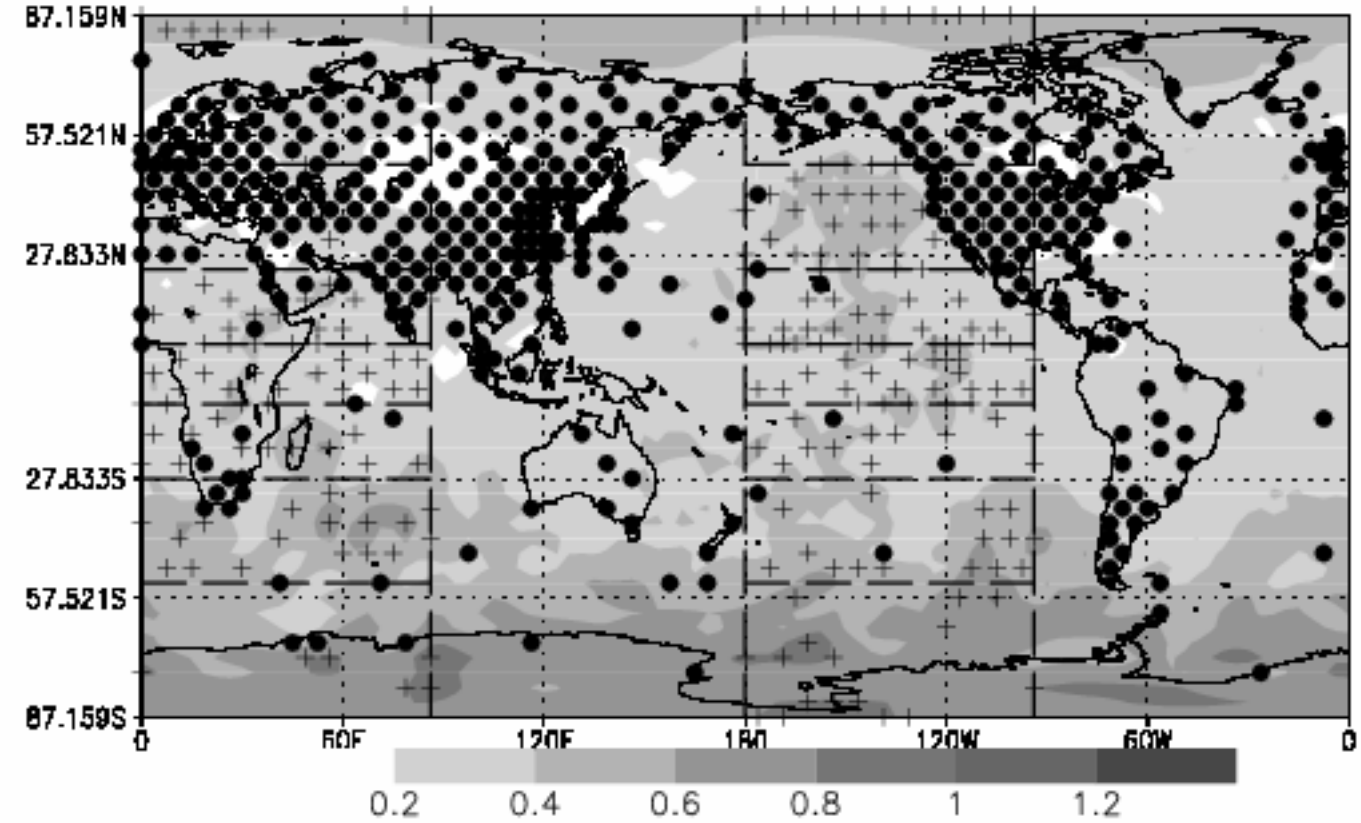
Note: 3D-Var and LETKF use the same locations in above strategies

- "Ideal" sampling

  → Locations with large background error obtained from the "truth".

Constraint: two adaptive observations have to be at least two grid points apart to avoid cluster of adaptive observations

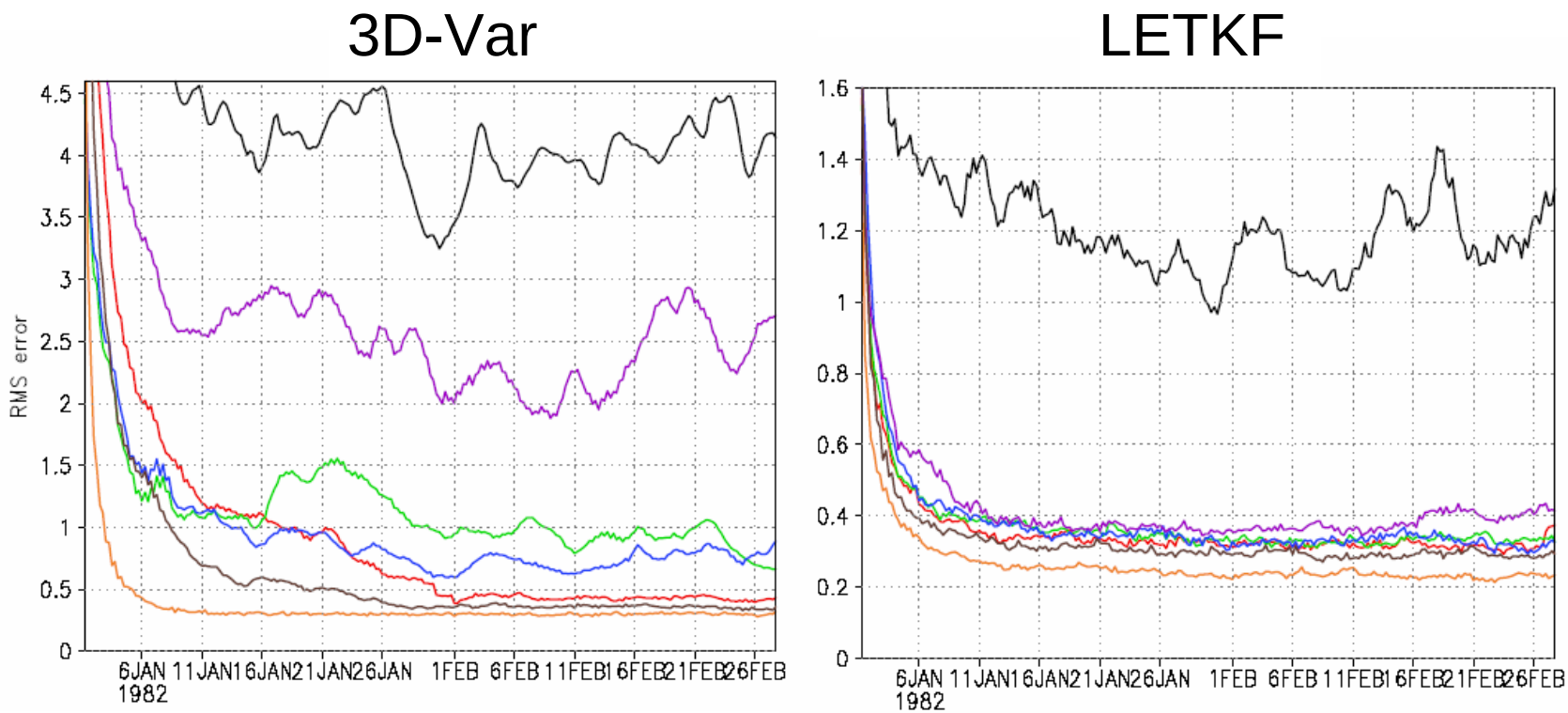
10% adaptive observation (+) distribution from ensemble spread (shaded area; unit: m/s) strategy of LETKF



- Total adaptive observation number: 10% of half global grid points.
- The number of adaptive observation in each latitude bands is proportional to the area of each band

500hPa zonal wind RMS error (10% adaptive obs)

Rawinsonde; climatology; uniform; random; ensemble spread; “ideal”; 100%

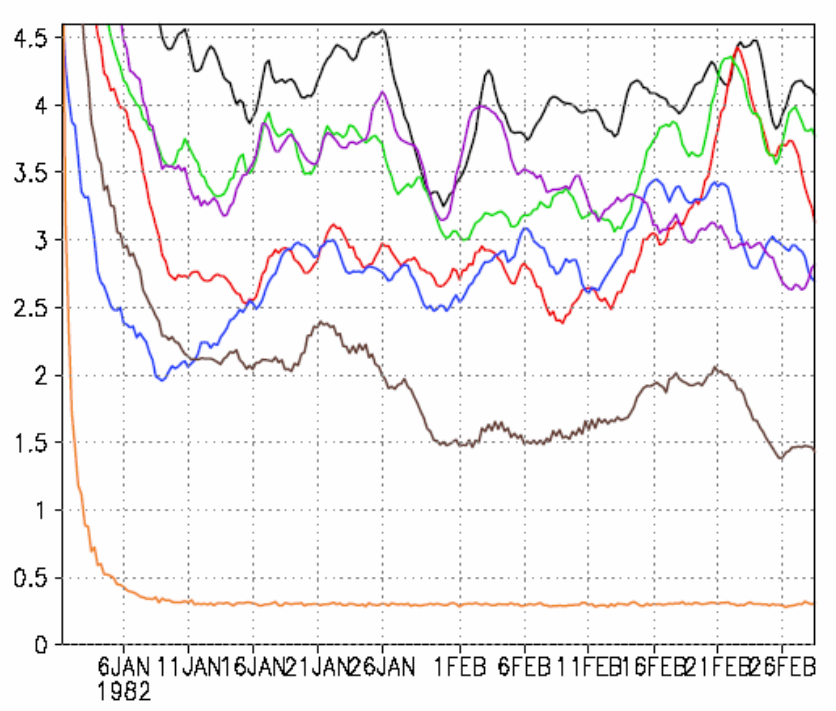


- The analysis accuracy is significantly improved for both 3D-Var and LETKF.
- Ensemble spread strategy gets best result among operational possible strategies.
- 3D-Var is more sensitive to adaptive strategies than LETKF.

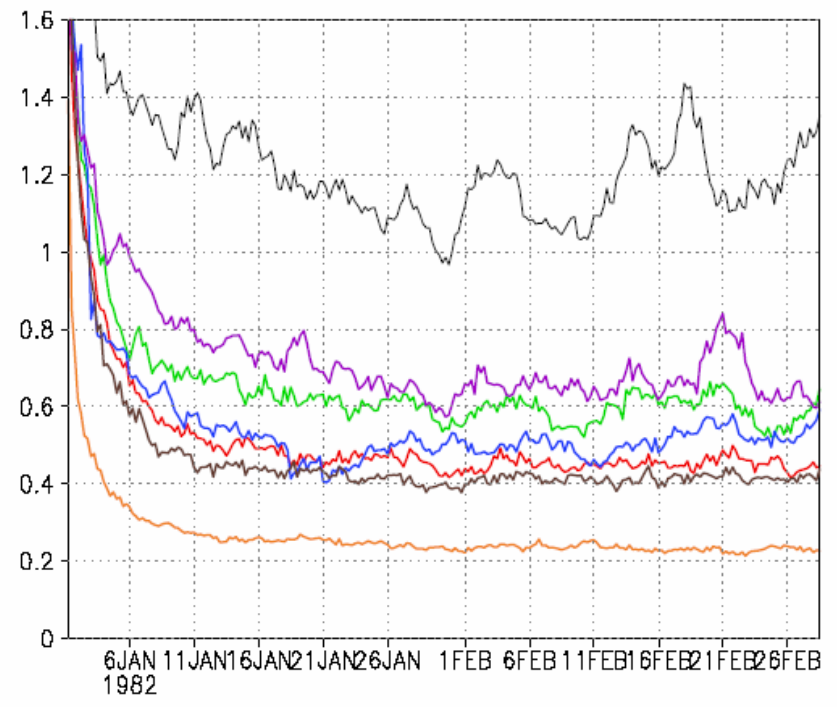
500hPa zonal wind RMS error (2% adaptive obs)

Rawinsonde; climatology; uniform; random; ensemble spread; “ideal”; 100%

3D-Var



LETKF



- With fewer (2%) adaptive observations, ensemble spread sampling strategy outperforms the other methods in LETKF
- For 3D-Var, 2% adaptive observations are not enough to make significant improvement with any method

Conclusions on adaptive observations

- With the constraint that two adaptive observations have to be at least two grid points apart, ensemble spread is close to optimal.

- 3D-Var is more sensitive to adaptive observation locations than the LETKF with 10% adaptive wind observations.

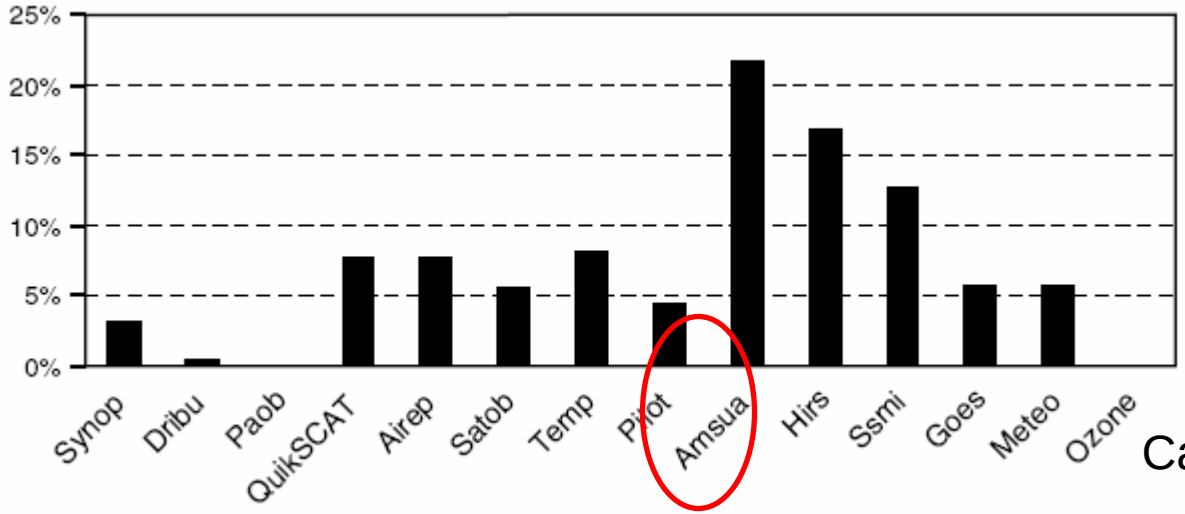
- 3D-Var is as effective as LETKF with 10% simulated DWL observations, but not as effective as LETKF with 2% simulated DWL adaptive observations.

Outline

- New applications of the LETKF
  - Adaptive observations
  - Analysis sensitivity and information content
  - Observation impact
  - Assimilation of moisture
- Summary

Analysis sensitivity and information content

Information content for different observation types



Cardinali et al., 2004

- Analysis sensitivity: how sensitive is the analysis to the observations.
- Information content (trace of analysis sensitivity): qualitatively reflects the importance of different type observations.

MY GOAL: calculate analysis sensitivity within the LETKF and study the properties of this quantity

Calculation of analysis sensitivity within the LETKF

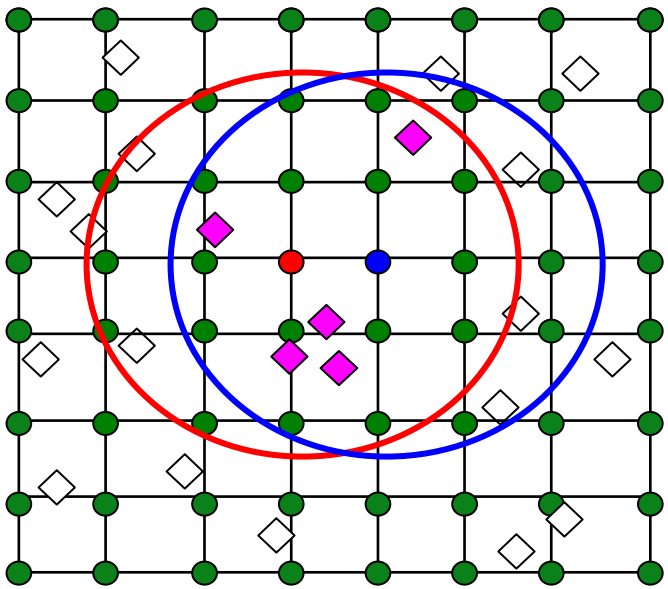
Definition: analysis sensitivity is the diagonal value of the influence matrix:

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}} = \mathbf{K}^T \mathbf{H}^T = \mathbf{R}^{-1}\mathbf{H}\mathbf{P}^a\mathbf{H}^T$$

- In 4D-Var (Cardinali et al., 2004), it requires an approximation to get $\mathbf{P}^a$
- $\mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1}$ is explicitly calculated in each local patch in the LETKF.

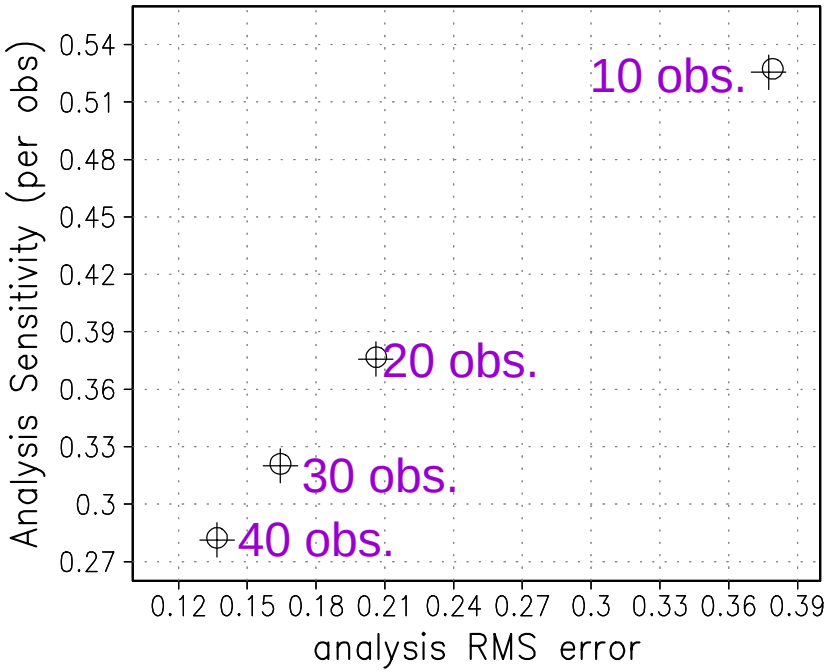
? The analysis sensitivity is different with respect to the same observation in different local patches.

⇒ Averaged the analysis sensitivity over the different local patches.



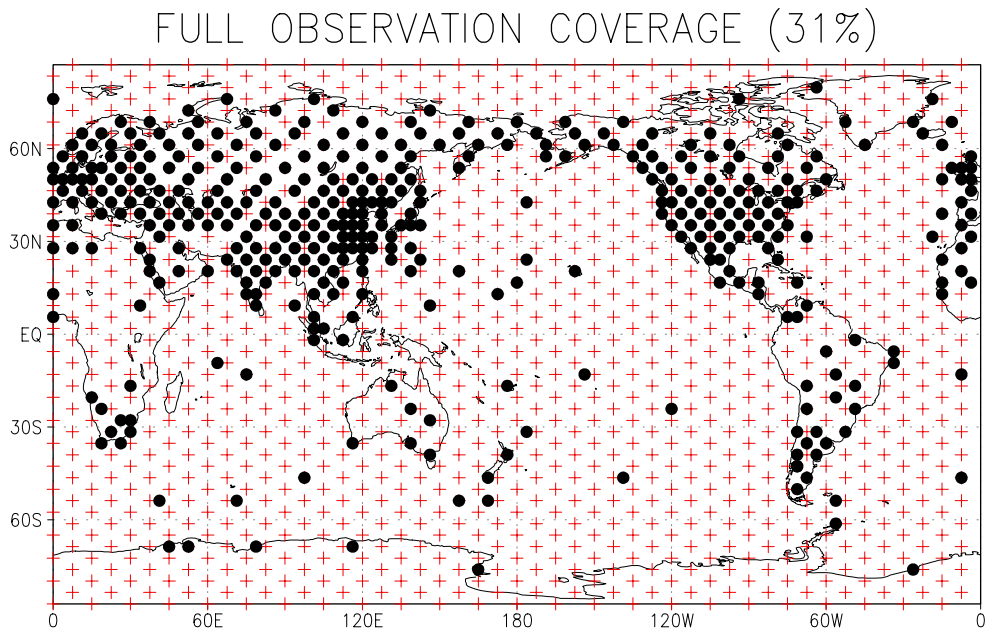
Verification of analysis sensitivity calculation method within the LETKF with Lorenz-40 variable model

The analysis sensitivity per obs.
from LETKF (⚐) & global ETKF (+)



- LETKF gives same results as global ETKF without averaging
- It decreases with the increasing of observation coverage, increases with the magnitude of the analysis error.

Simulated experiments with SPEEDY



Data denial experiments:

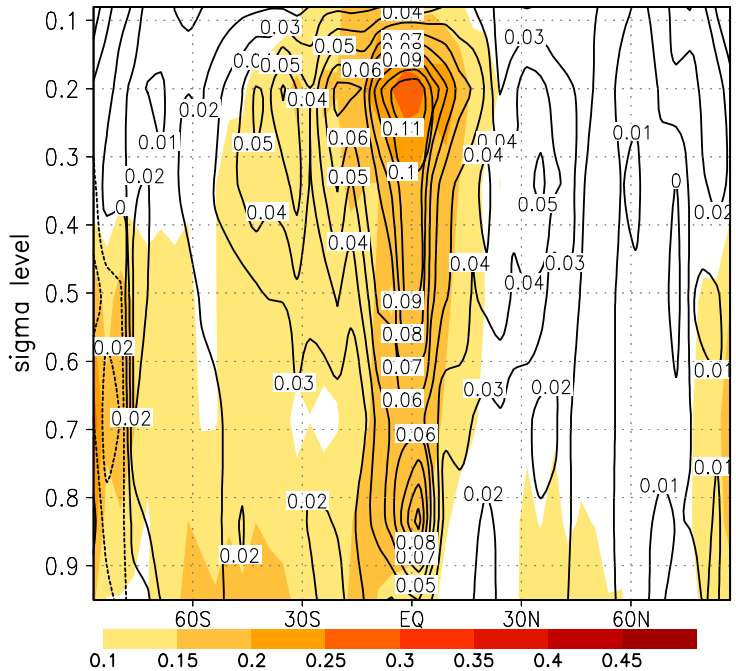
Control run: full coverage for all dynamical variables

Sensitivity experiment: u is not observed in locations with red +

• Compare information content (the trace of analysis sensitivity) of zonal wind at locations with red + from control run to the RMS error difference between sensitivity experiment and *control experiment*.

Information content (control, shaded) vs. RMSE difference (data-denial experiments, contour)

RMSE (u, *sensitivity-control*) & info-content (u)



- Information content qualitatively reflects the actual observation impact from data-denial experiments.

Summary of analysis sensitivity

- ✓ The calculation of analysis sensitivity (like Cardinali et al., 2004) needs no approximation and can be calculated along with the data assimilation in the LETKF.
- ✓ The trace of analysis sensitivity qualitatively reflects the actual observation impact from much more expensive data-denial experiments.

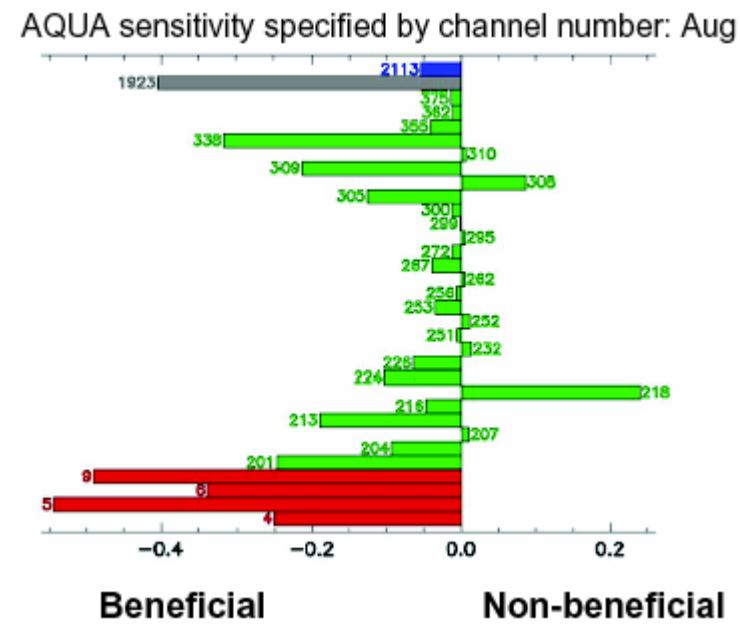
Summary of analysis sensitivity

- ✓ The calculation of analysis sensitivity (like Cardinali et al., 2004) needs no approximation and can be calculated along with the data assimilation in the LETKF.
- ✓ The trace of analysis sensitivity qualitatively reflects the actual observation impact from much more expensive data-denial experiments.
- × But analysis sensitivity cannot give quantitative estimation of observation impact, and cannot detect the observations that have poor quality

Outline

- New applications of the LETKF
  - Adaptive observations
  - Analysis sensitivity and information content
  - Observation impact (like Langland and Baker, 04, Zhu and Gelaro, 07, but without adjoint)
  - Assimilation of moisture
- Summary

Motivation of observation impact



AIRS shortwave 4.180 µm

AIRS shortwave 4.474 µm

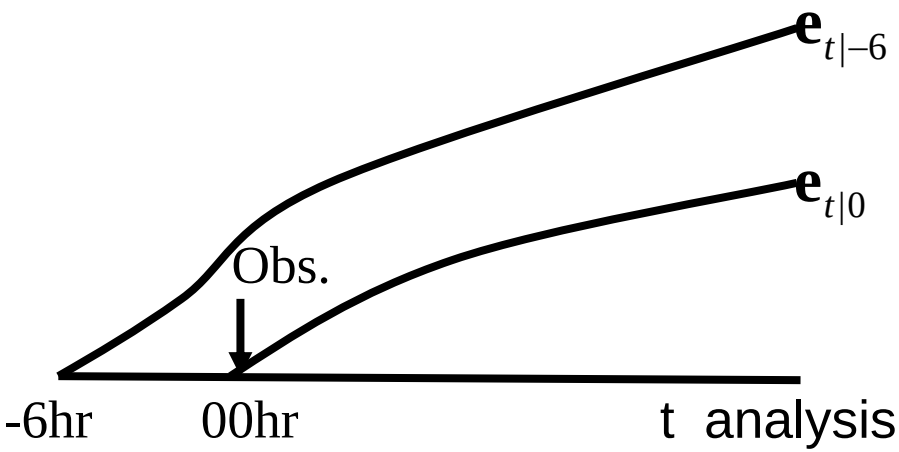
AIRS longwave 14-13 µm

AMSU/A

- The adjoint method proposed by Langland and Baker (2004) quantifies the reduction in forecast error for each individual satellite channel.
- The adjoint method detects the observations which make the forecast worse.
- With the adjoint of GSI, Zhu and Gelaro (2007) carries similar observation impact as Langland and Baker (2004).

MY GOAL: propose an ensemble sensitivity method to calculate observation impact without using adjoint model.

Schematic of the observation impact on the forecast error



$$\mathbf{e}_{t|-6} = \overline{\mathbf{x}}^{f}_{t|-6} - \overline{\mathbf{x}}^{a}_{t}$$

$$\mathbf{e}_{t|0} = \overline{\mathbf{x}}^{f}_{t|0} - \overline{\mathbf{x}}^{a}_{t}$$

(Adapted from Langland and Baker, 2004)

The only difference between $\mathbf{e}_{t|0}$ and $\mathbf{e}_{t|-6}$ is the assimilation of observations at 00hr.

- Cost function:

$$J = \frac{1}{2}\left(\mathbf{e}^{T}_{t|0}\mathbf{e}_{t|0} - \mathbf{e}^{T}_{t|-6}\mathbf{e}_{t|-6}\right)$$

The ensemble sensitivity method

Euclidian cost function: $$J = \frac{1}{2}(\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T \mathbf{e}_{t|-6}) \qquad \mathbf{v}_0 = \mathbf{y}_0^o - h(\overline{\mathbf{x}}_{0|-6}^b)$$

Cost function as function of obs. Increments: $$J = \left\langle \mathbf{v}_0, \frac{\partial J}{\partial \mathbf{v}_0} \right\rangle$$

The ensemble sensitivity method

Euclidian cost function: $$J = \frac{1}{2}(\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T \mathbf{e}_{t|-6}) \qquad \mathbf{v}_0 = \mathbf{y}_0^o - h(\overline{\mathbf{x}}_{0|-6}^b)$$

Cost function as function of obs. Increments: $$J = \left\langle \mathbf{v}_0, \frac{\partial J}{\partial \mathbf{v}_0} \right\rangle$$

We derived the sensitivity of forecast error change to the assimilated observations:

$$\frac{\partial J}{\partial \mathbf{v}_0} = \left[\tilde{\mathbf{K}}_0^T \mathbf{X}_{t|-6}^{fT} \right] \left[\mathbf{e}_{t|-6} + \mathbf{X}_{t|-6}^f \tilde{\mathbf{K}}_0 \mathbf{v}_0 \right]$$

The cost function as function of assimilated observations:

$$J = \left\langle \mathbf{v}_0, \frac{\partial J}{\partial \mathbf{v}_0} \right\rangle = \sum_{i=1}^{n} \left(\frac{\partial J}{\partial v_0^i} \cdot v_0^i \right)$$

Experimental design

- Model: Lorenz-40 variable model (Lorenz and Emanuel, 1998)

- Full observation coverage

- Three experiments:

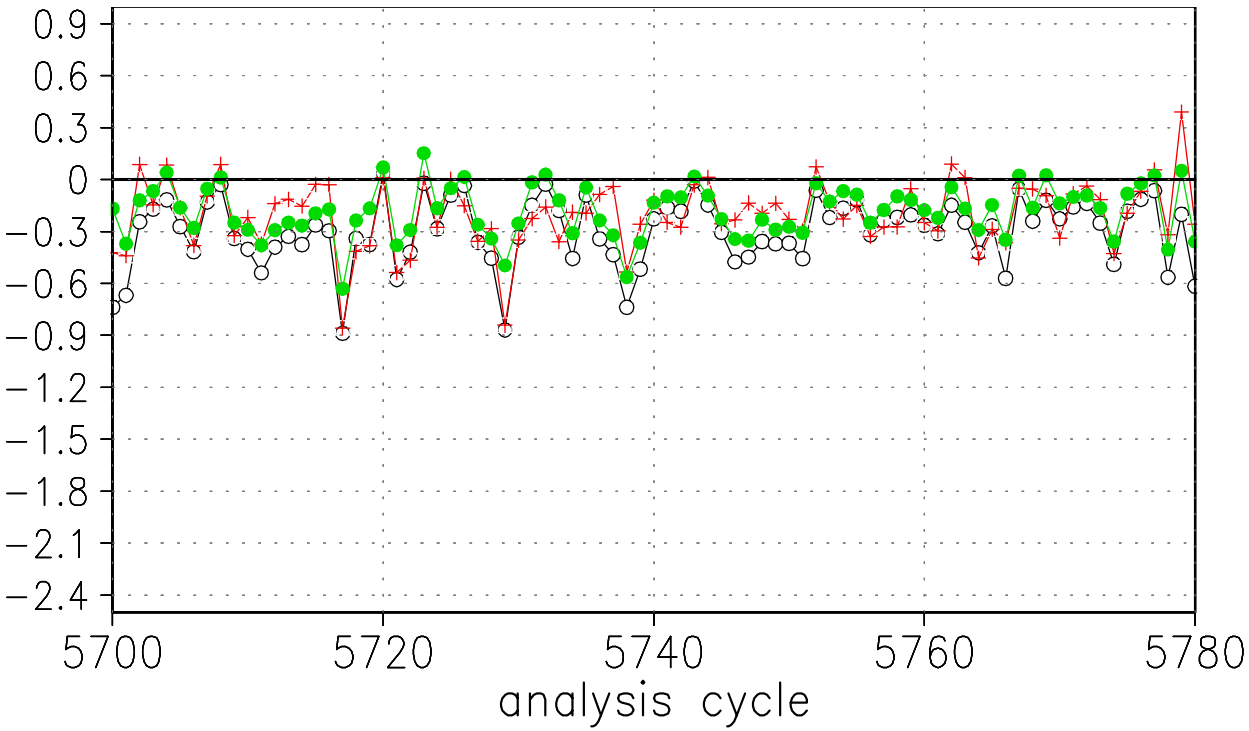
- **Normal**: observation error is 0.2 at every observation location.

- **Larger random error**: the observation error SD at 11th grid point is 0.8, but still assume 0.2 in the data assimilation.

- **Bias**: the observation at 11th observation location has a bias equal to 0.5.

Observation impact comparison between adjoint method (LB) and ensemble sensitivity method in normal case

Adjoint method (red), ensemble method (green) and actual forecast error reduction (black)

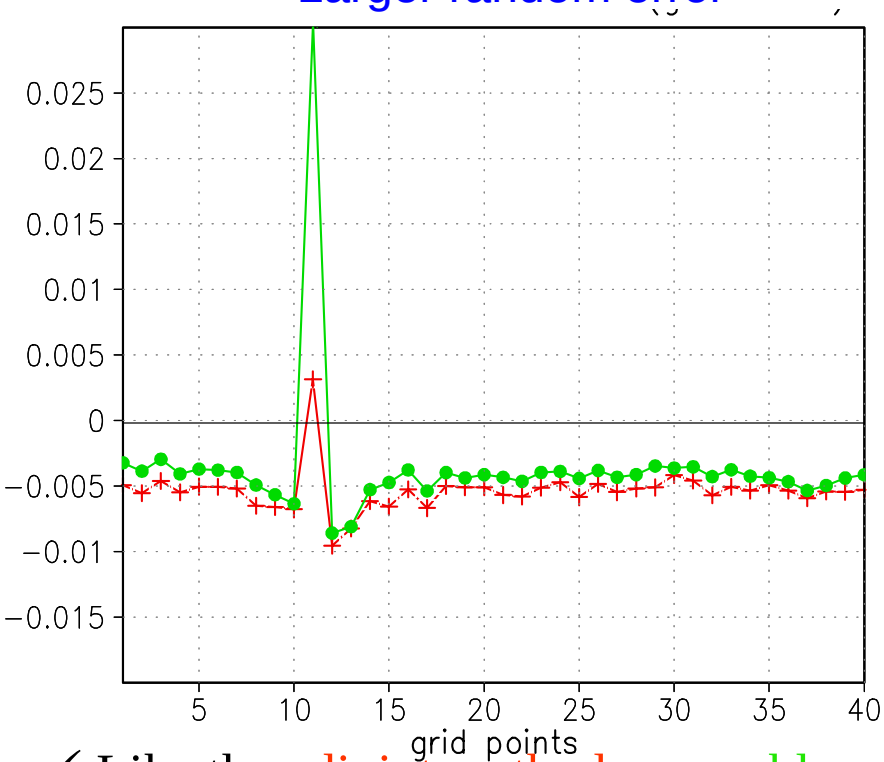


- ✓ The ensemble sensitivity method gives results similar to the adjoint method
- ✓ Both reflect most of the actual observation impact (black) in the forecast.

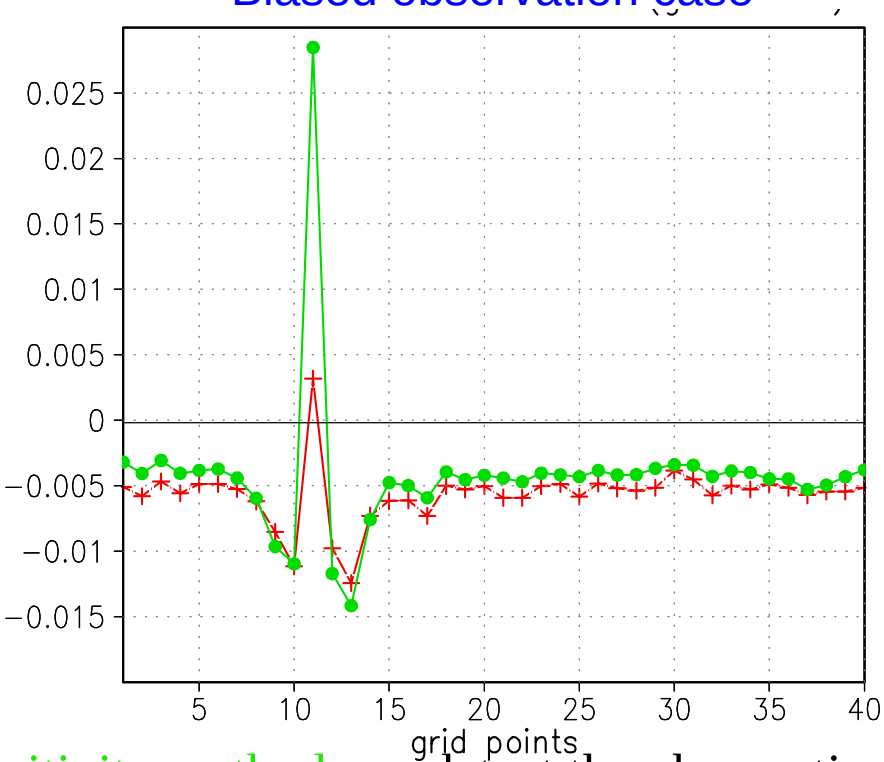
Ability to detect the poor quality observation

Observation impact from LB (red) and from ensemble sensitivity method (green)

Larger random error



Biased observation case



✓ Like the adjoint method, ensemble sensitivity method can detect the observation poor quality (11th observation location)

✓ The ensemble sensitivity method has a stronger signal when the observation has negative impact on the forecast.

Conclusions of observation impact

- Derived a formula to calculate the observation impact based on the ensemble without using the adjoint model which usually is not available.

- The results based on Lorenz-40 variable model show that ensemble sensitivity method without using adjoint model gives results similar to adjoint method .

- Like the adjoint method, ensemble sensitivity method can detect the observation which either has larger random error or has bias. Under such conditions, the ensemble sensitivity method has stronger signal.

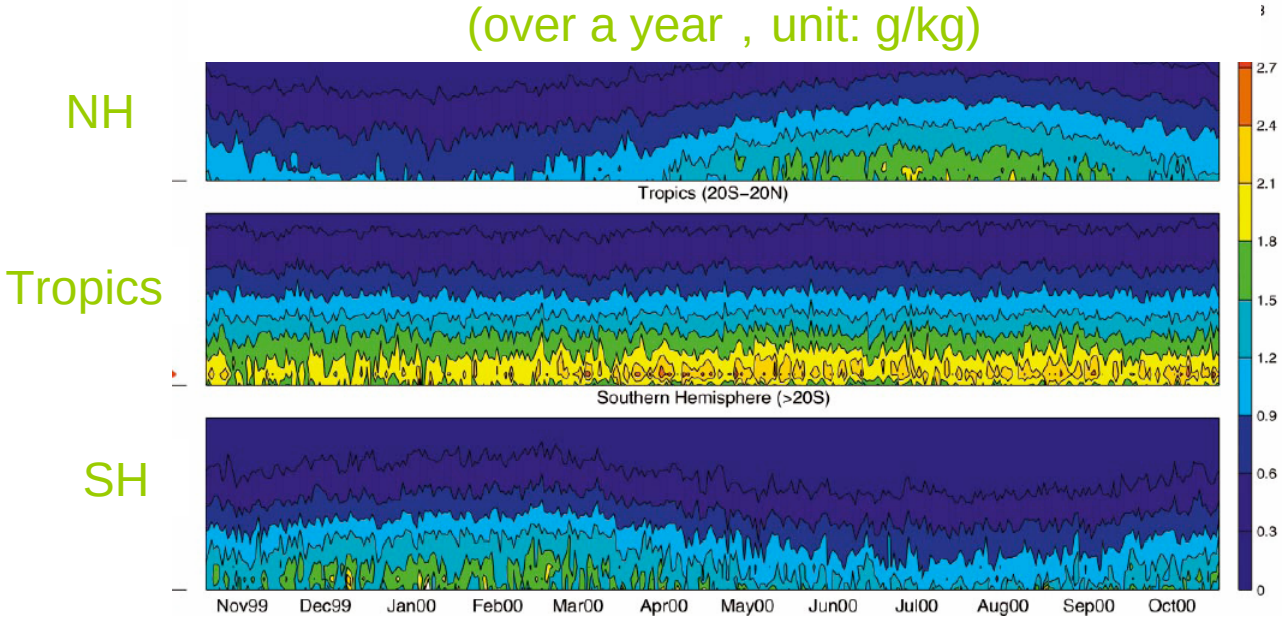
- This provides a powerful tool to check the quality of the observations.

Outline

- New applications of the LETKF
  - Adaptive observations
  - Analysis sensitivity
  - Observation impact and information content
  - Assimilation of moisture
- Summary

Motivation of assimilation of moisture

Specific humidity obs. minus background
(over a year , unit: g/kg)



Dee and Da Silva, 2003

Daily RMS statistics of rawinsonde observed-minus-background mixing ratio residuals during the period 1 Nov 1999—31 Oct 2000, produced by fvDAS.

- The specific humidity error varies abruptly and is very non-Gaussian

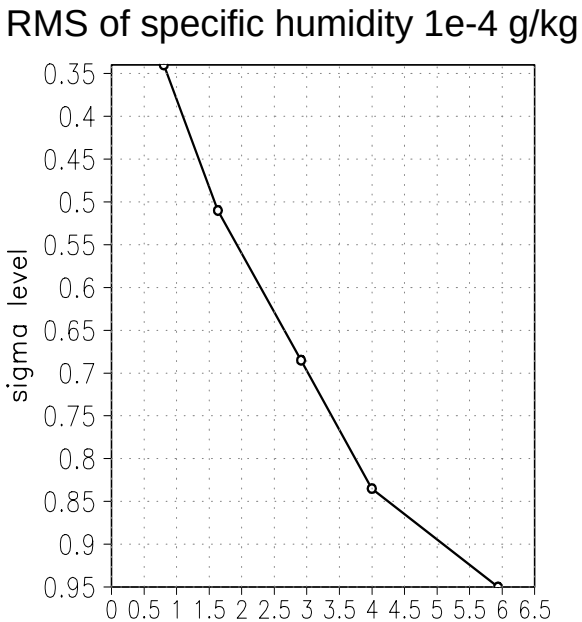
MY GOALS:

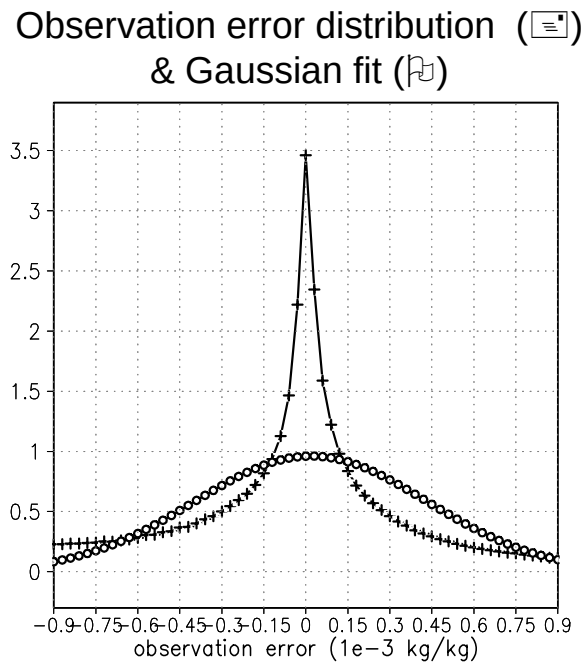
- Assimilate humidity with the LETKF
- Multivariate assimilation of humidity.
- The impact of different choices of humidity variables on the analysis results

Assimilation humidity on perfect model

- Control run:
  - The humidity observations are not assimilated.
- Humidity run
  - *Uni-variate* assimilation of humidity.
  - *Multivariate* assimilation of humidity.
- Observations
  - u, v, T, Ps observations are nature run with Gaussian random errors
  - Humidity observations are created with Gaussian errors in ln(q)

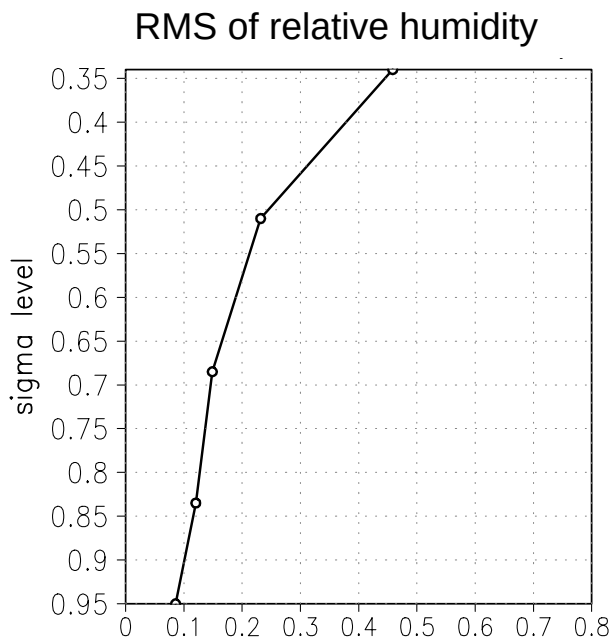
Choice of humidity variable (1): specific humidity (q)

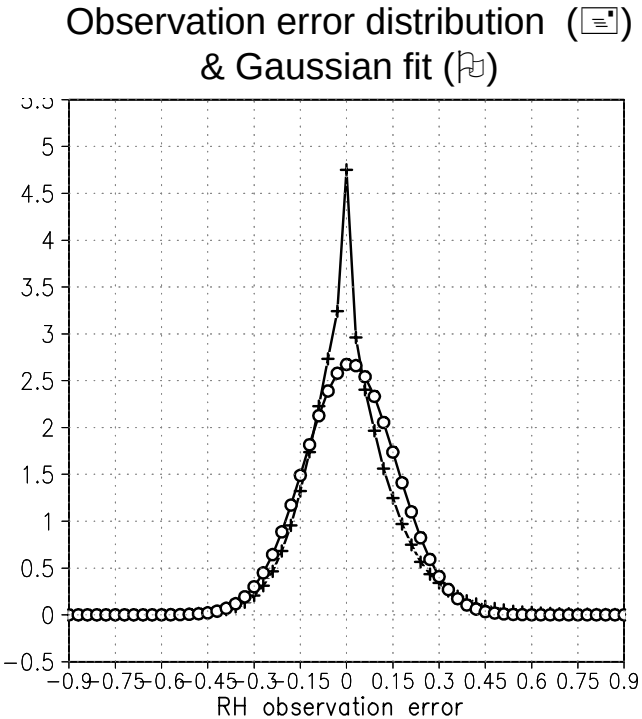




- **Specific humidity (q^o)**: exponential of the ln(q) observation.
- Observation error changes with vertical levels abruptly.
- Specific humidity observation errors are far from Gaussian.

Choice of humidity variable (2): relative humidity (RH)





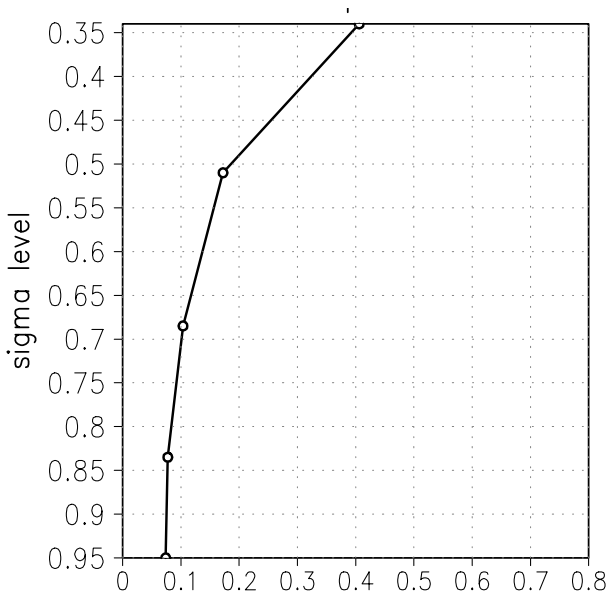
➲ **Relative humidity**: $\dfrac{q^o}{q^{sat}(T^o, P^o)}$

➢ It has a more Gaussian error distribution than specific humidity

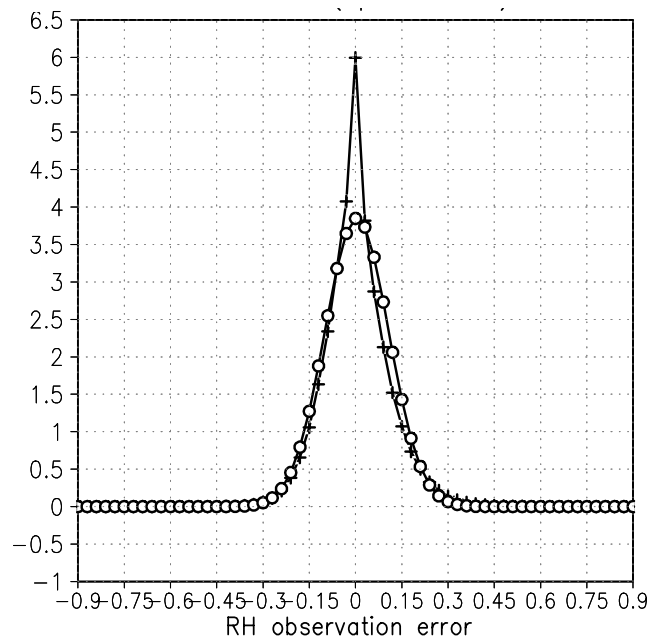
➢But, RH errors have high correlation with temperature observation errors.

Choice of humidity variable (3): pseudo relative humidity (pseudo-RH, Dee and Da Silva, 2003)

RMS of pseudo-RH



Observation error distribution (⌨)
& Gaussian fit (⚐)

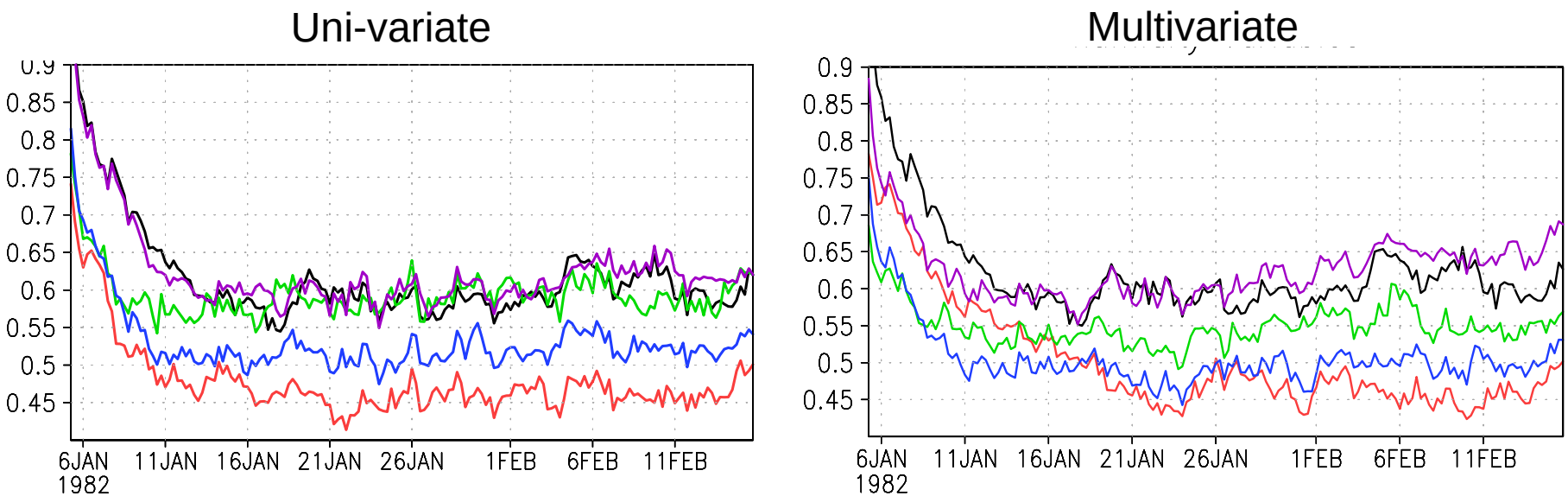


➲ Pseudo-RH: $\dfrac{q^o}{q^{sat}(T^b, P^b)}$

- It has a similar error distribution (more Gaussian) as the relative humidity
- Pseudo-RH errors do not correlate with the temperature observation errors.

700hPa specific humidity analysis RMS error (10^{-4} g/kg)

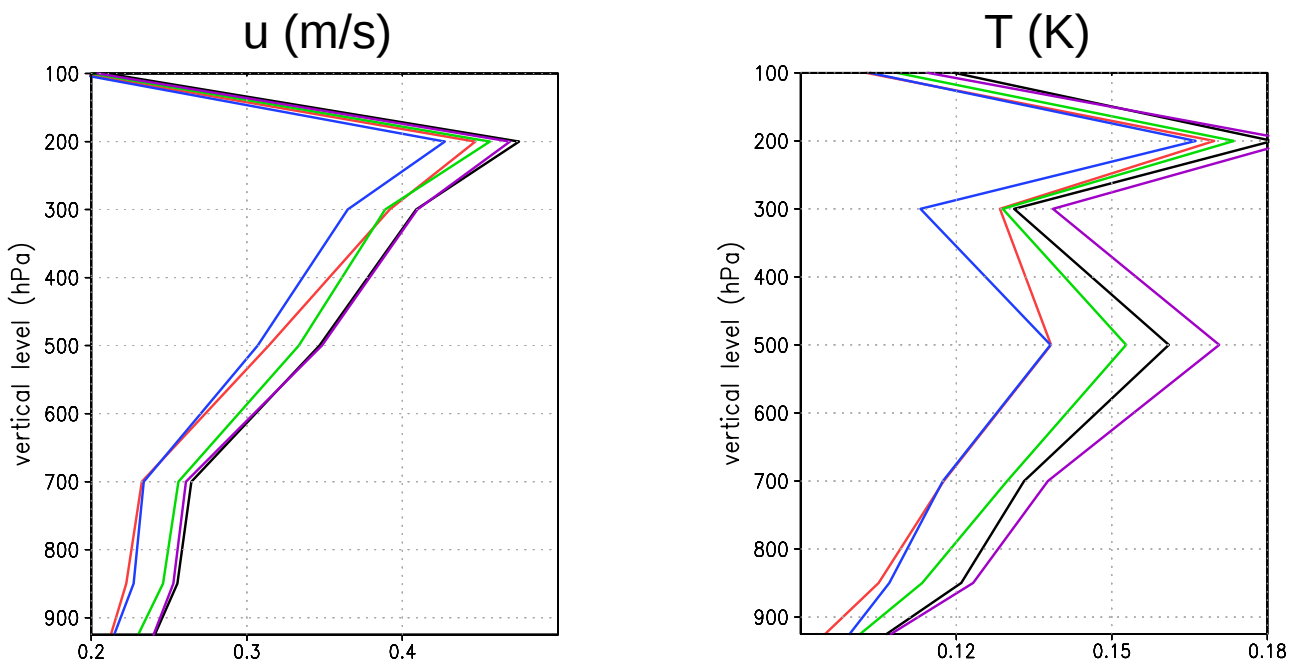
Black: control; Purple: RH; Green: q; Blue: pseudo-RH; Red: ln(q)



- Multivariate assimilation is better than uni-variate assimilation
- ln(q) is a standard for the other choices to attain in our OSSEs.
- Pseudo-RH gives the best result among the other choices of humidity variables.

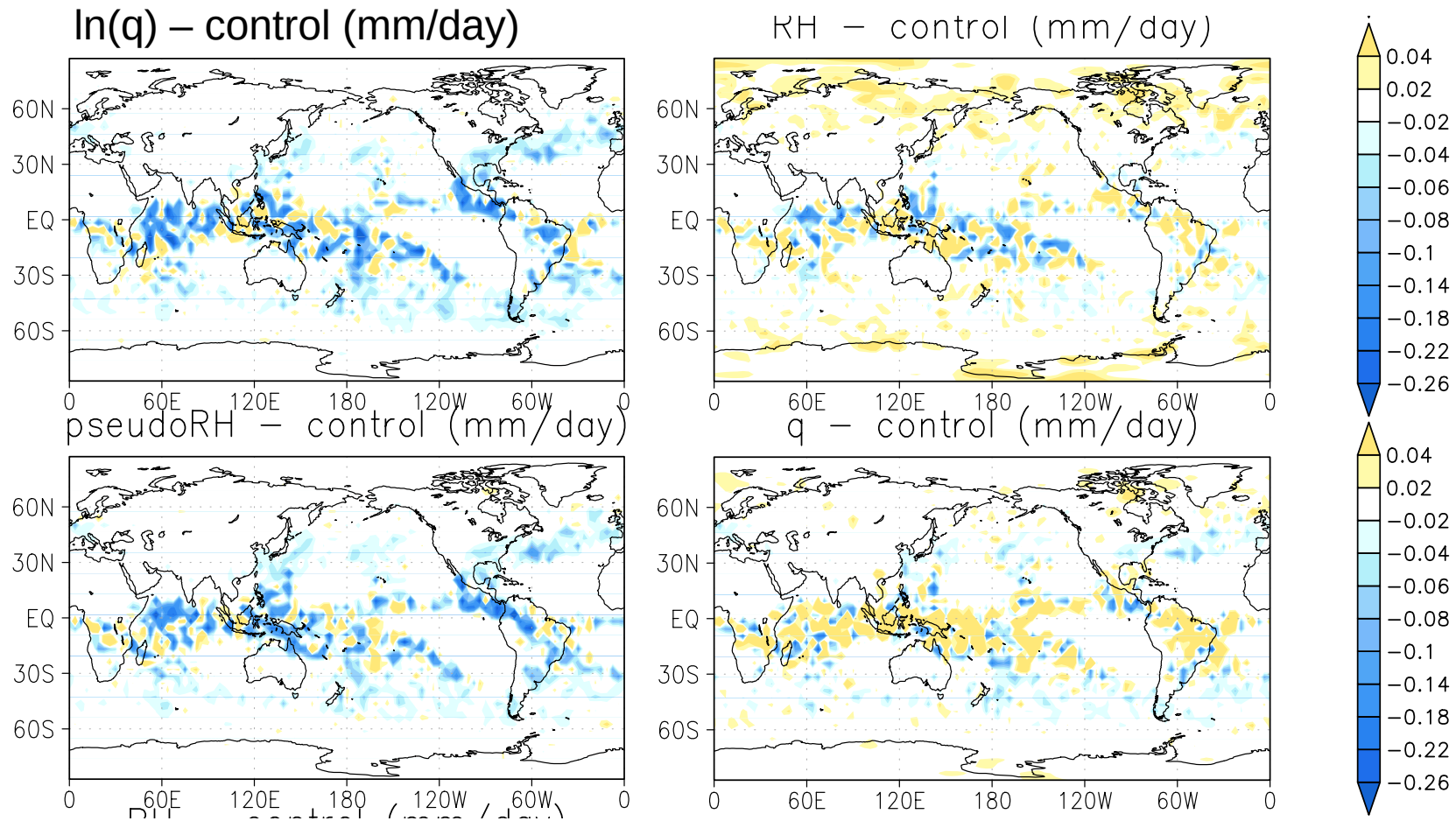
Time average analysis RMS error comparison in multivariate experiments

Black: control, Red: ln(q); Blue: pseudo-RH; Green: q; Purple: RH;



- Assimilation of humidity improves the analysis accuracy for wind and temperature.
- With pseudo-RH, the analysis results of zonal wind and temperature are even better than with ln(q) assimilation

Time average total precipitation 6-hour forecast RMS error comparison between *humidity* (multivariate) and *control* run



- Precipitation forecast becomes worse in most of the regions in q and RH experiments.
- The 6-hour precipitation forecast accuracy is improved in most of the area in pseudo-RH and ln(q) experiments.

Assimilation of real humidity data

AIRS humidity retrievals (from Chris Barnet) on NCEP GFS with LETKF

Control Run: Non-radiance data (Szunyogh, et al. 2007) plus AIRS temperature retrievals (Li et al., 2007).

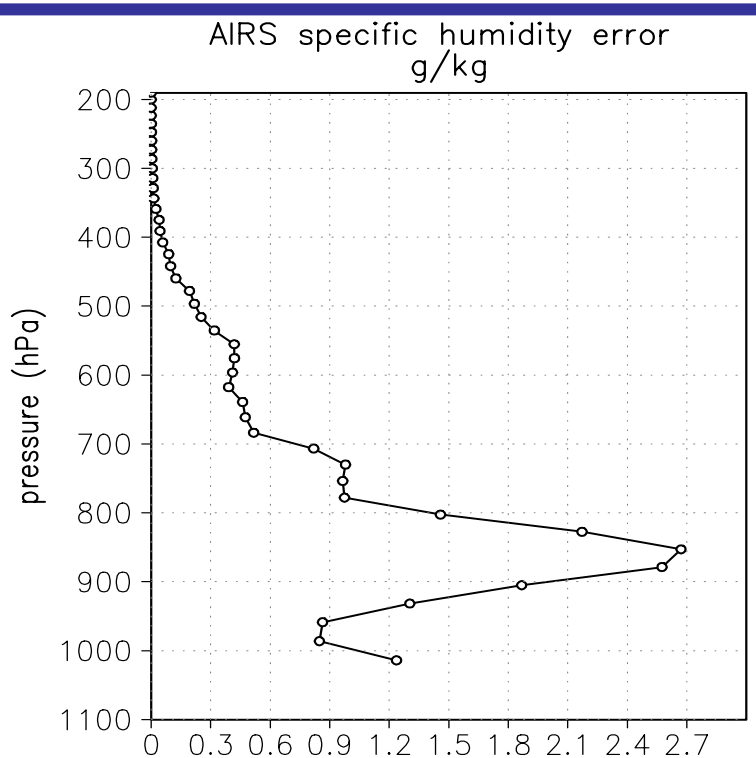
Humidity Run:

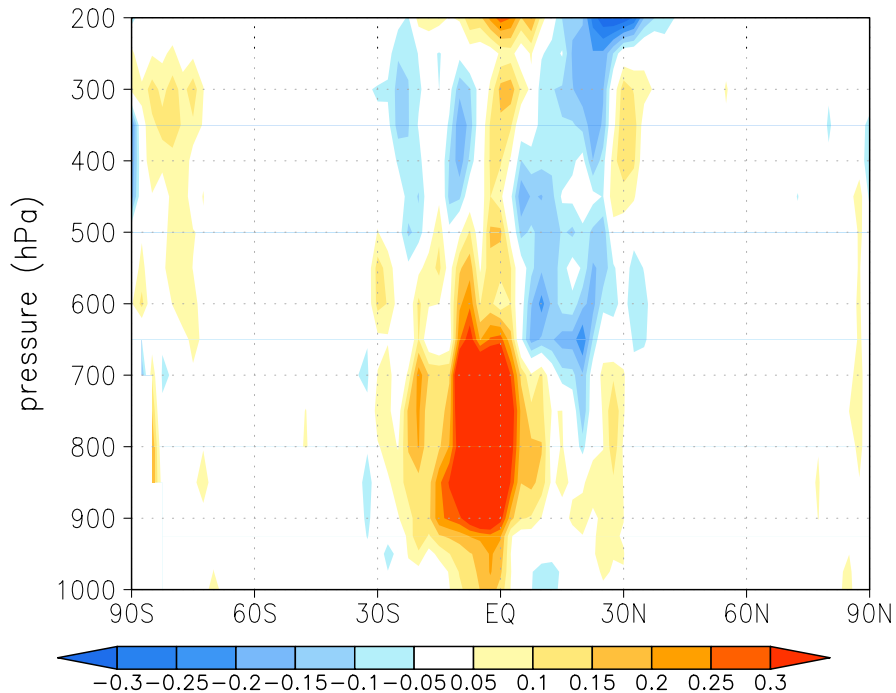
- Added AIRS specific humidity retrievals
  - within 30ºS and 30ºN (below 200hPa)
  - globe (below 200hPa)

- Multivariate assimilation of humidity variable (for the first time) with the choice of q and pseudo-RH.

Verification: Operational NCEP analysis at T254L64, assimilating all operational observations.

Specific humidity observation error standard deviation (L) & relative humidity RMS error difference between humidity run and control run (R)



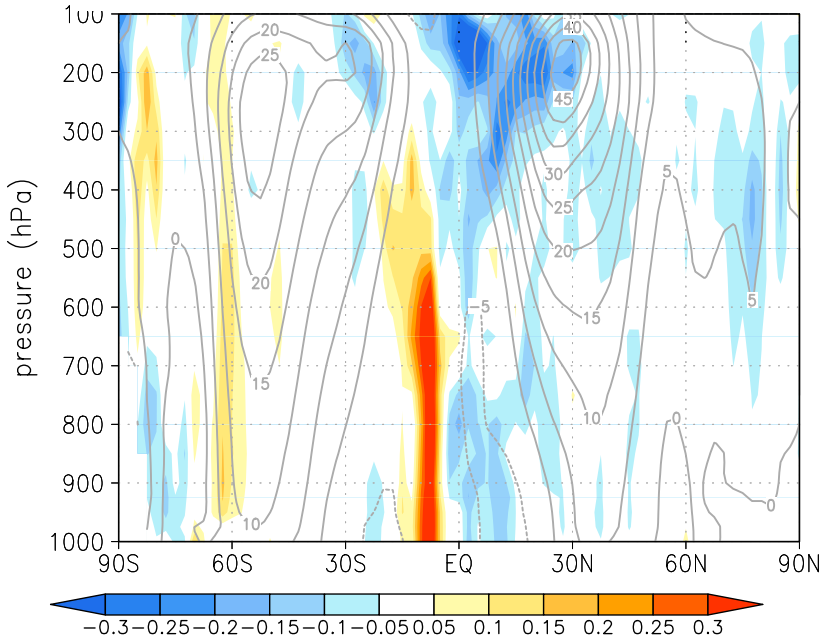


- The impact mainly concentrates over the region where we assimilate specific humidity retrievals.
- Positive impact on the upper tropics, negative impact on the lower tropics where errors are large.

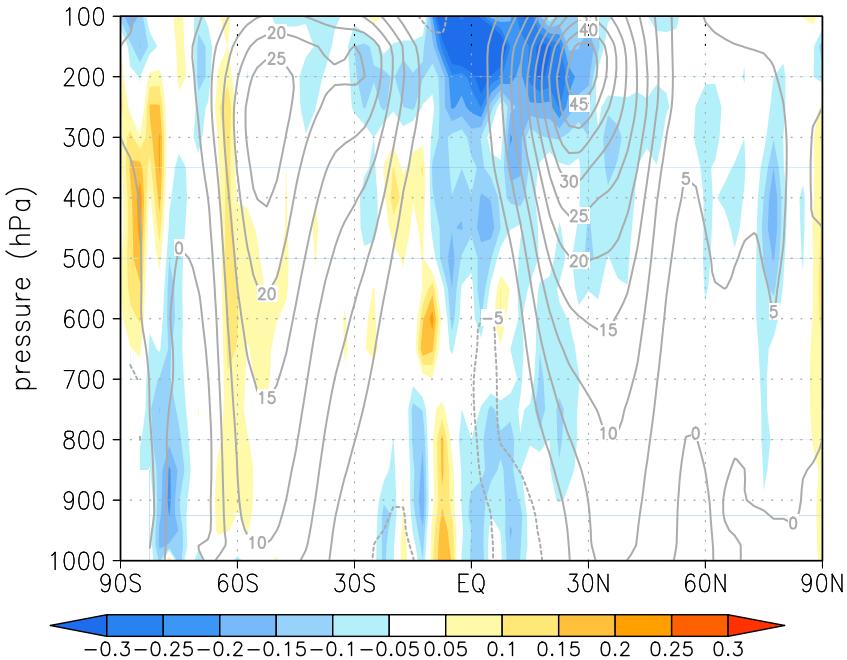
Zonal wind RMS error difference (shaded) between humidity run and control run & time average zonal wind field (contour)

Humidity obs between 30°S and 30°N

(assimilating q)



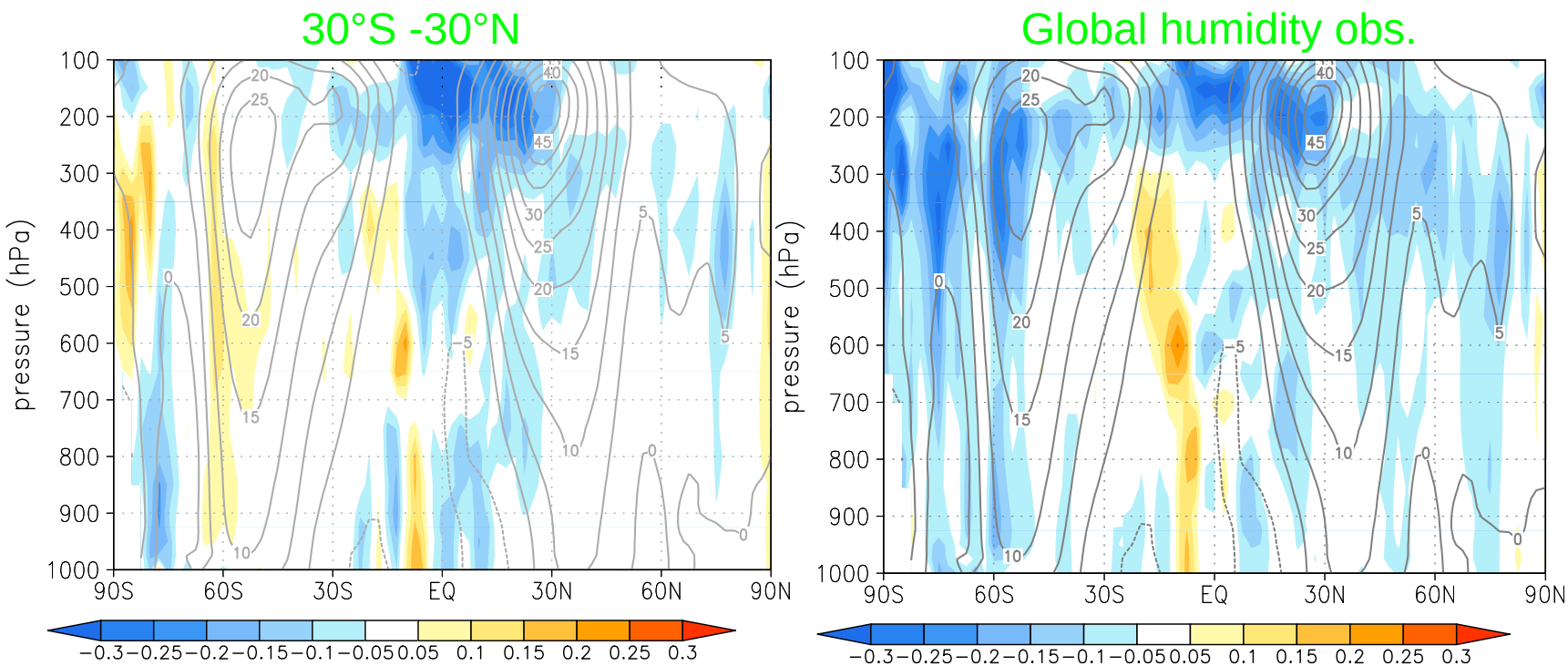
(assimilating pseudo-RH)



- Strong positive impact (blue color) in wind in the upper tropics and the NH
- with the choice of pseudo-RH, the analysis accuracy is further improved

Zonal wind RMS error difference (shaded) between humidity run and control run & time average zonal wind field (contour)

(assimilating pseudo-RH)



- Adding more data in high latitudes further improve the analysis accuracy.

Conclusions of humidity assimilation

- EnKF reflects time-changing error statistics and automatically couples the errors of the dynamical variables including humidity.

- Based on perfect model assimilation:
  - Pseudo-RH gives better results than RH and specific humidity.
  - Multivariate experiments are better than uni-variate experiments.

- Based on real data assimilation:
  - Preliminary result with multivariate assimilation of AIRS specific humidity retrievals shows positive impact on wind analysis results.
  - Pseudo-RH gives better results than the choice of specific humidity
  - Global assimilation of pseudo-humidity improves the winds throughout the globe.

Outline

- New applications of the LETKF
  - Adaptive observations
  - Analysis sensitivity
  - Observation impact
  - Assimilation of Moisture
- Summary

Summary

Application	Main results
Adaptive observations	▪ LETKF-based ensemble spread strategy (with constraint) is very effective
Analysis sensitivity	▪ Proposed a method to calculate observation information content within the LETKF without approximation
Observation impact study	▪ Derived an ensemble sensitivity method without using adjoint model: very powerful tool to check quality of observation
Assimilation of moisture	▪ Pseudo-RH gives better results than q ▪ First multivariate assimilation of real humidity observations with improved winds