CSCAMM-DAS13 – Lecture 3

- Effective assimilation of precipitation (Guo-Yuan Lien, Kalnay, Miyoshi, Tellus, 2013)
- Forecast sensitivity to observations (Kalnay et al, 2012, Ota et al., 2013, Tellus, 2013)
  - "Proactive Quality Control"
- Application of forecast sensitivity to data assimilation (Yang and Kalnay, in progress, thanks to Enomoto).
- Estimation of surface fluxes as evolving parameters (Kang et al, 2011, 2012, JGR)

Guo-Yuan Lien, Shu-Chih Yang, Yoichiro Ota, T. Miyoshi, Ji-Sun Kang and Eugenia Kalnay

UMD Weather-Chaos Group: **Kayo Ide**, **Brian Hunt**, Ed Ott, and students (Guo-Yuan Lien, Yan Zhou, Adrienne Norwood, Erin Lynch, Yongjing Zhao, Daisuke Hotta, Travis Sluka)

Also: **Y Ota**, Juan Ruiz, C Danforth, M Peña, M Corazza, A. Carrassi

Effective Assimilation of Precipitation

(Guo-Yuan Lien, E. Kalnay and T Miyoshi)

- Assimilation of precipitation has been done by changing the moisture *Q* in order to make the model “rain as observed”.
- Successful during the assimilation: e.g. the North American Regional Reanalysis had perfect precipitation!
- However the model **forgets** about the changes soon after the assimilation stops!
- The model **will remember potential vorticity (PV).**
- EnKF should modify PV efficiently, since the analysis weights will be larger for an ensemble member that is raining more correctly, because it has a better PV.

- However, 5 years ago, we had tried assimilating precipitation observations in a LETKF-SPEEDY model simulation but the results were POOR!
- Big problem: precipitation is not Gaussian.
- We tried a Gaussian transformation of precipitation and it worked!

How do we transform precipitation y to a Gaussian y_{transf}?

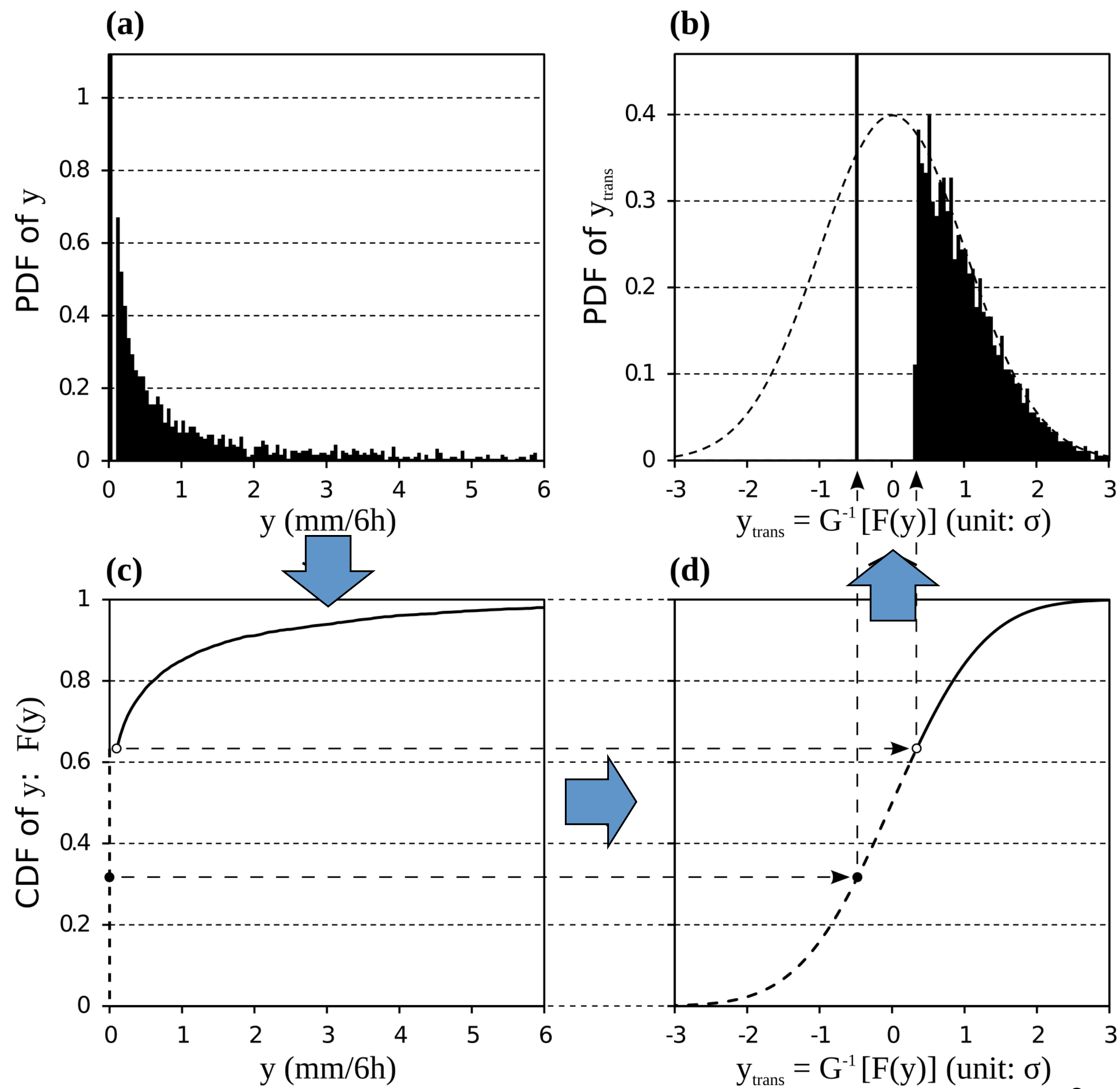
Start with pdf of y=rain at every grid point.

"No rain" is like a delta function that we cannot transform.

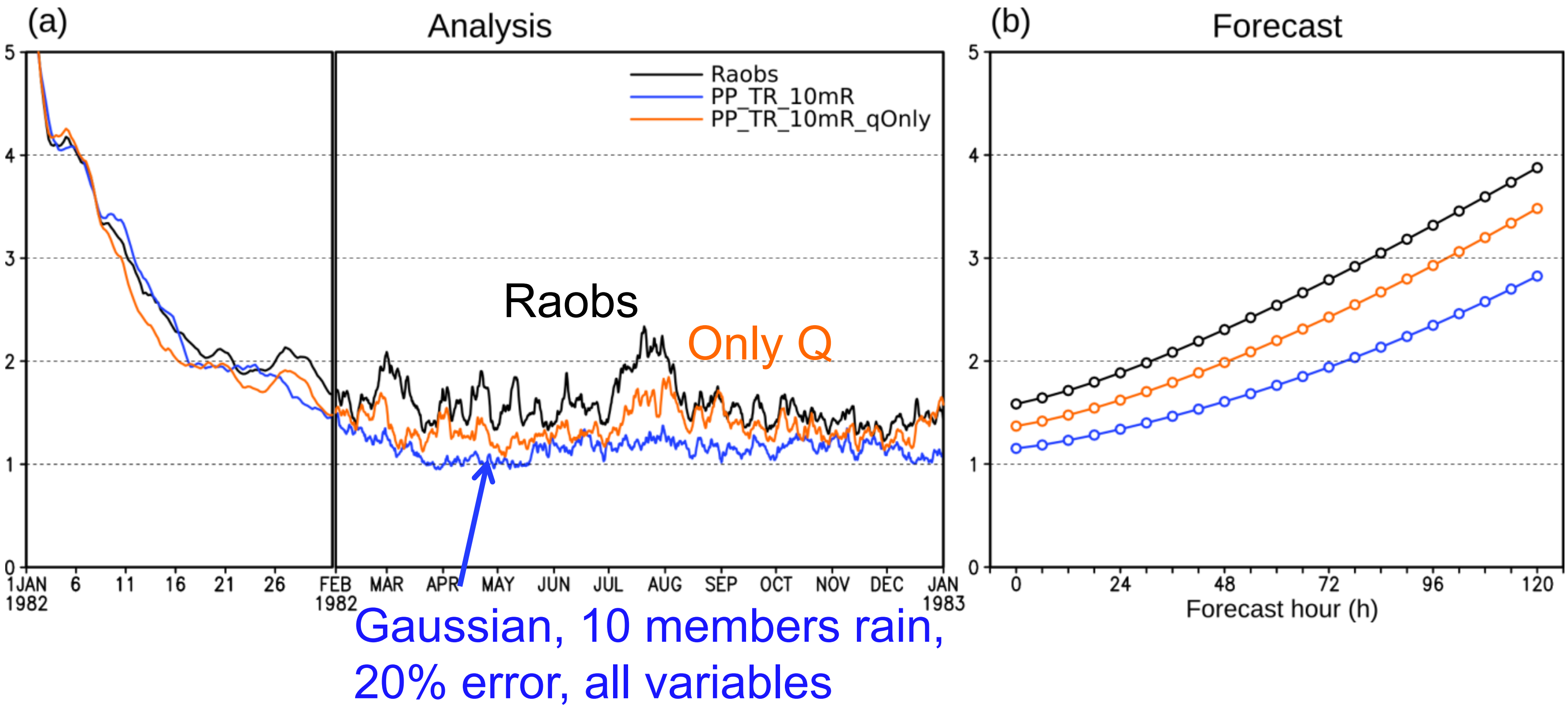
We assign all "no rain" to the **median** of the no rain CDF.

We found this works as well as more complicated procedures.

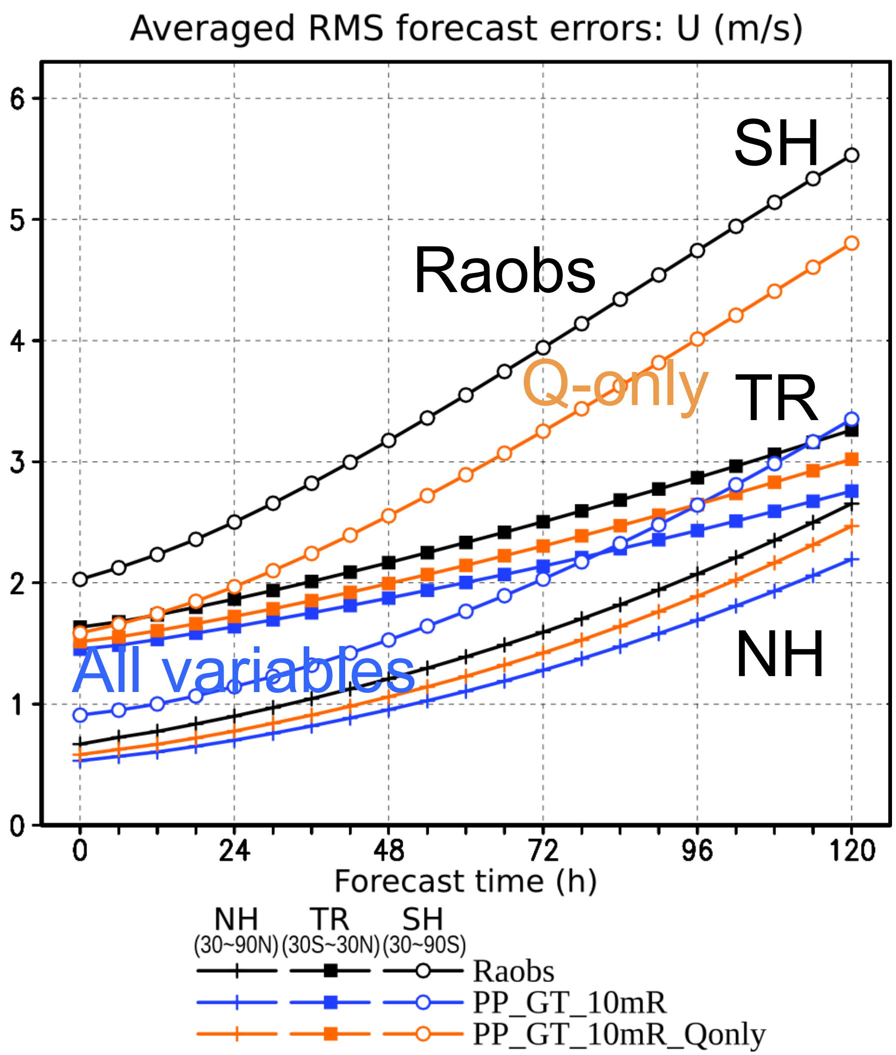
It allows to assimilate both rain and no rain.



$$G^{-1}(x) = \sqrt{2}\, erf^{-1}(2x-1)$$

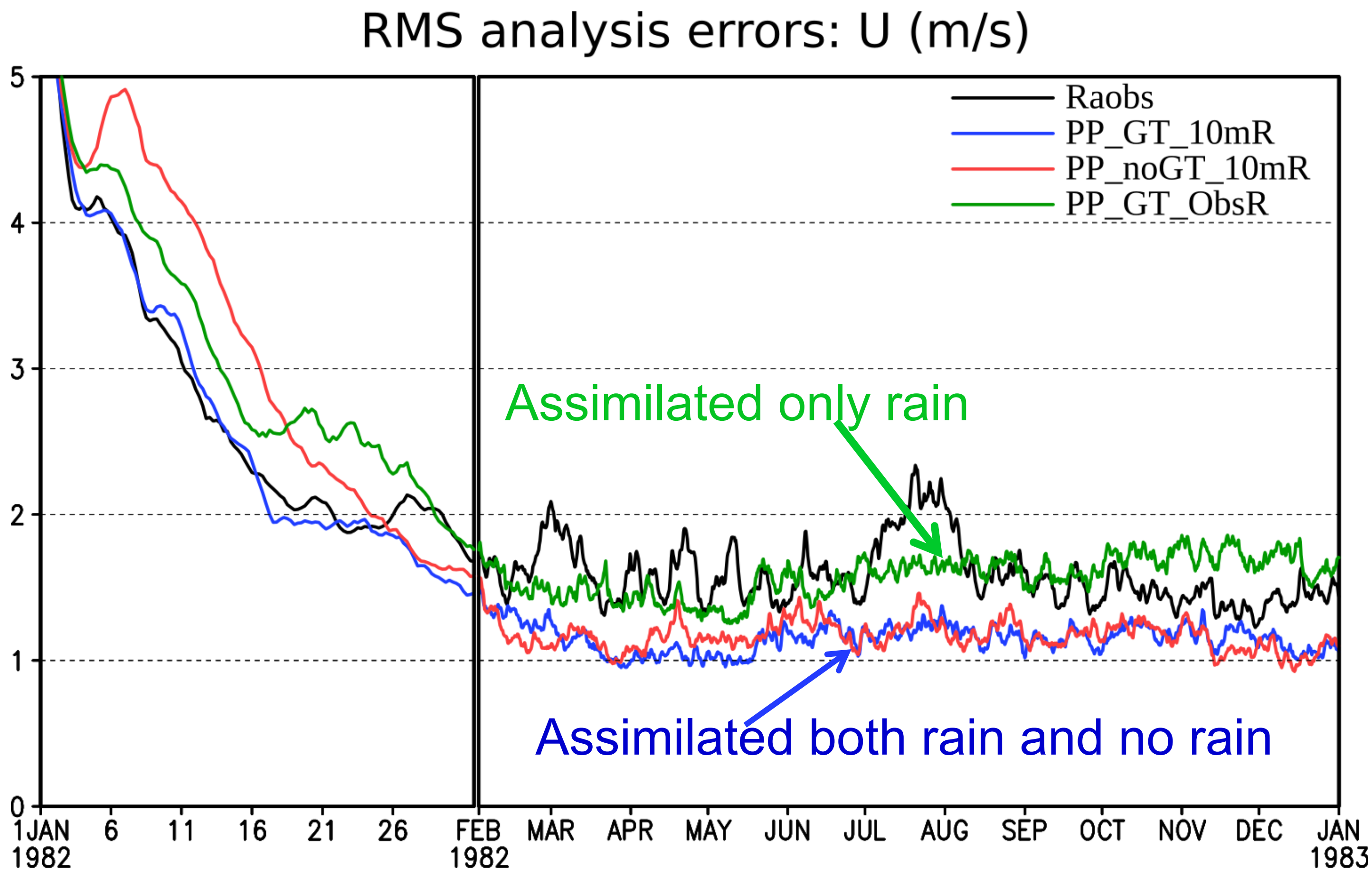


- Main result: with at least 10 ensemble members raining in order to assimilate an obs, updating all variables (including vorticity), with Gaussian transform, and rather accurate observations (20% errors), the analyses and forecasts are much improved!
- Updating only Q is much less effective.
- **The 5-day forecasts maintain the advantage.**

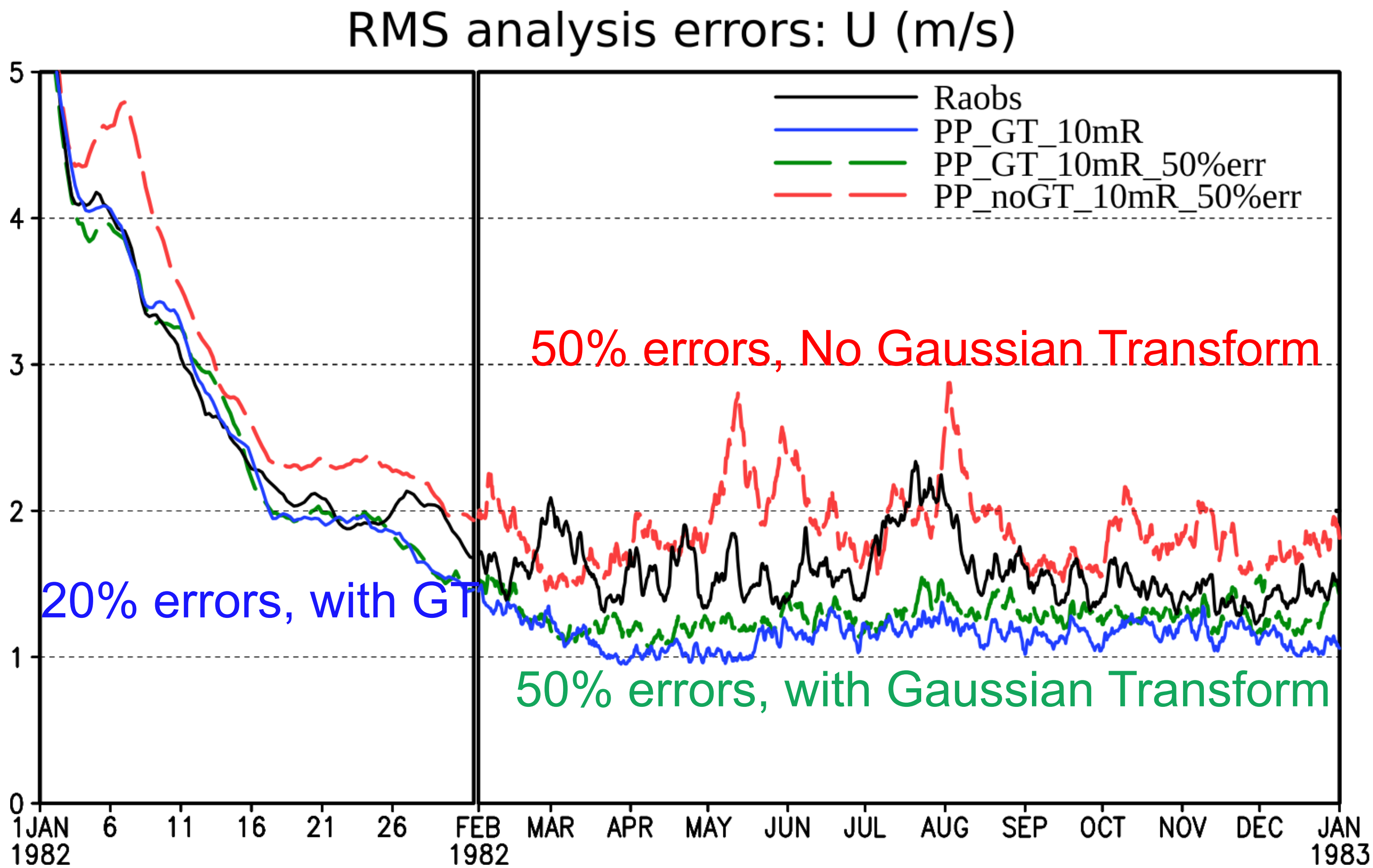


One year of 5-day forecasts

The model remembers the impact of pp assimilation in the SH, NH and tropics!

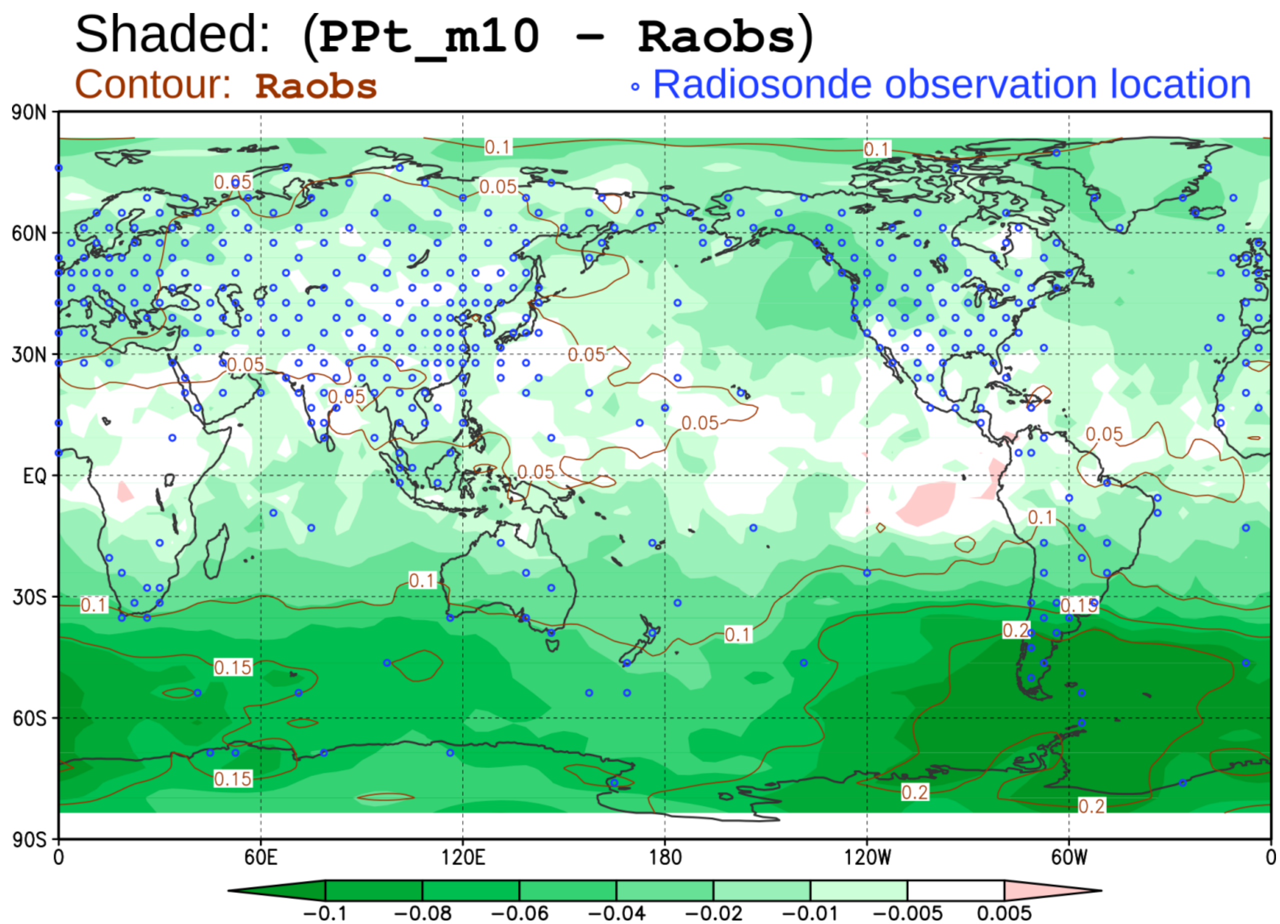


If we assimilate only rain the results are much worse!
We need to assimilate both rain and no rain!



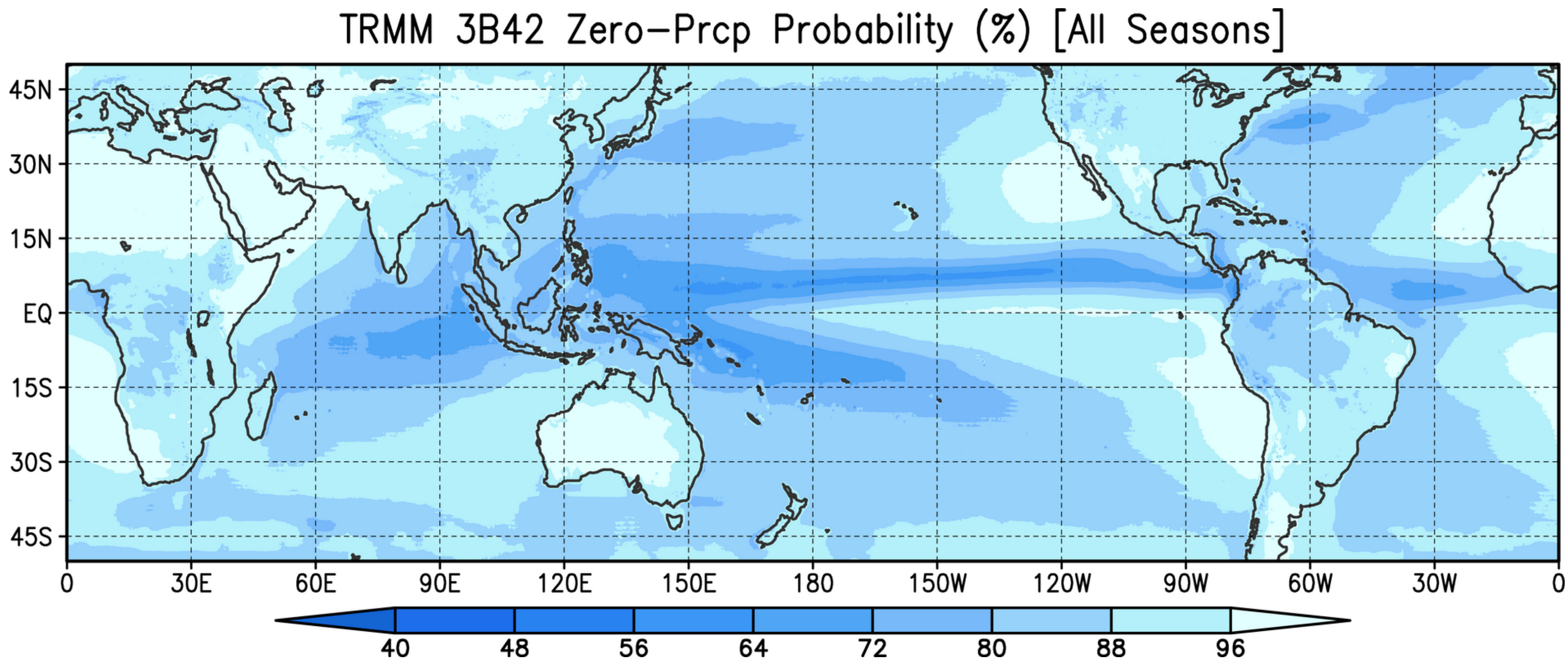
The impact of the Gaussian Transform is important with large observation errors (50% rather than 20%). The impact of GT50% is almost as good as GT20%.

Vorticity errors and corrections



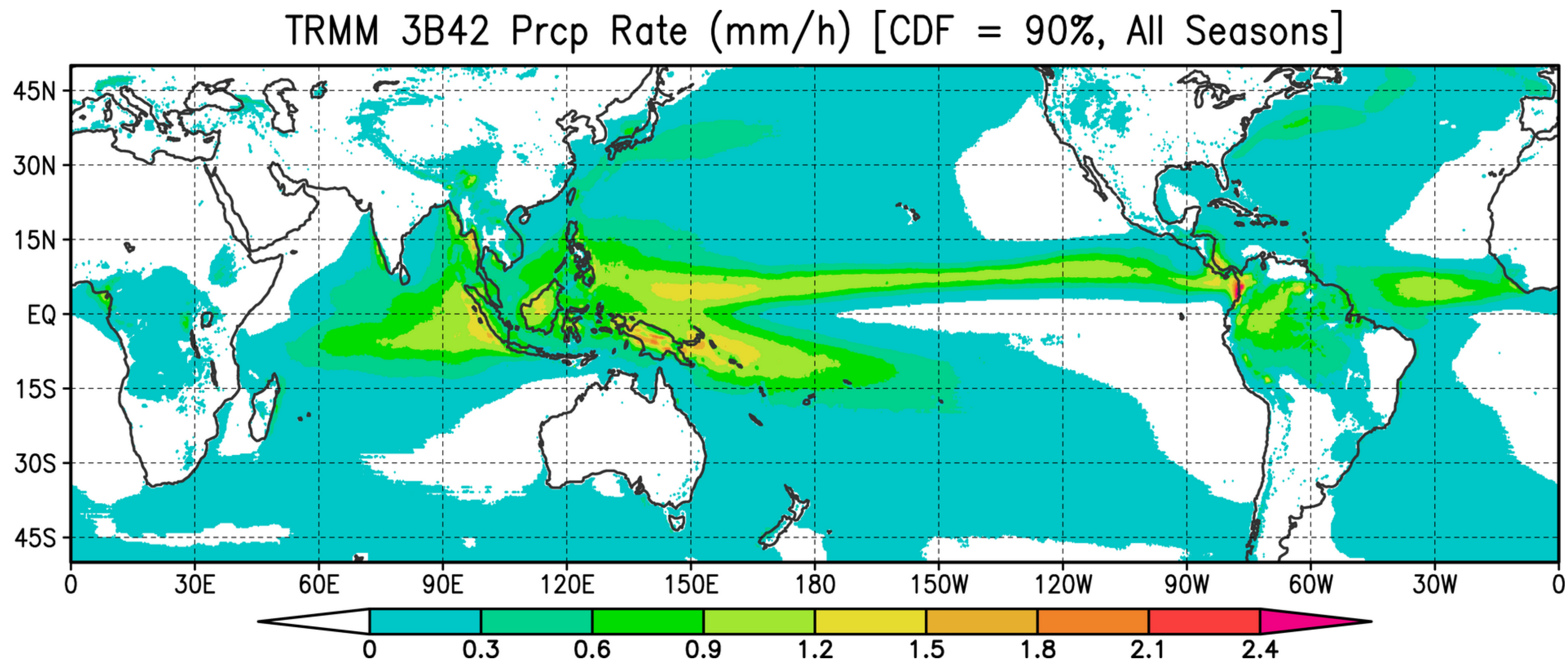
There is no vorticity information in the pp observations, but the LETKF clearly knows about the vorticity errors

How about real observations?
We will use TRMM/TMPA satellite estimates (from G. Huffman) with the NCEP GFS



TRMM/TMPA: 14 years of data, 50S-50N, 3hrs, 0.5 deg

TRMM/TMPA (data from G. Huffman)



TRMM/TMPA: 14 years of data, 50S-50N, 3hrs, 0.5 deg

Summary for assimilation of precipitation

- The model remembers potential vorticity (dynamics), does not remember moisture changes, or even temperature.
- For this reason, when using nudging, or variational assimilation of precipitation to change Q and T, the model “forgets” this information and returns to the original forecast.
- EnKF has a better chance to assimilate potential vorticity by giving higher weights to ensemble members with good precip.
- In addition, EnKF has the advantage of not requiring model linearization, a problem for variational systems.
- We found that EnKF with a Gaussian transformation of precipitation assimilates rain info and remembers it during the forecast.
- Requiring at least several ensemble forecasts to have Rain>0 allows the effective assimilation of both rain and no rain.

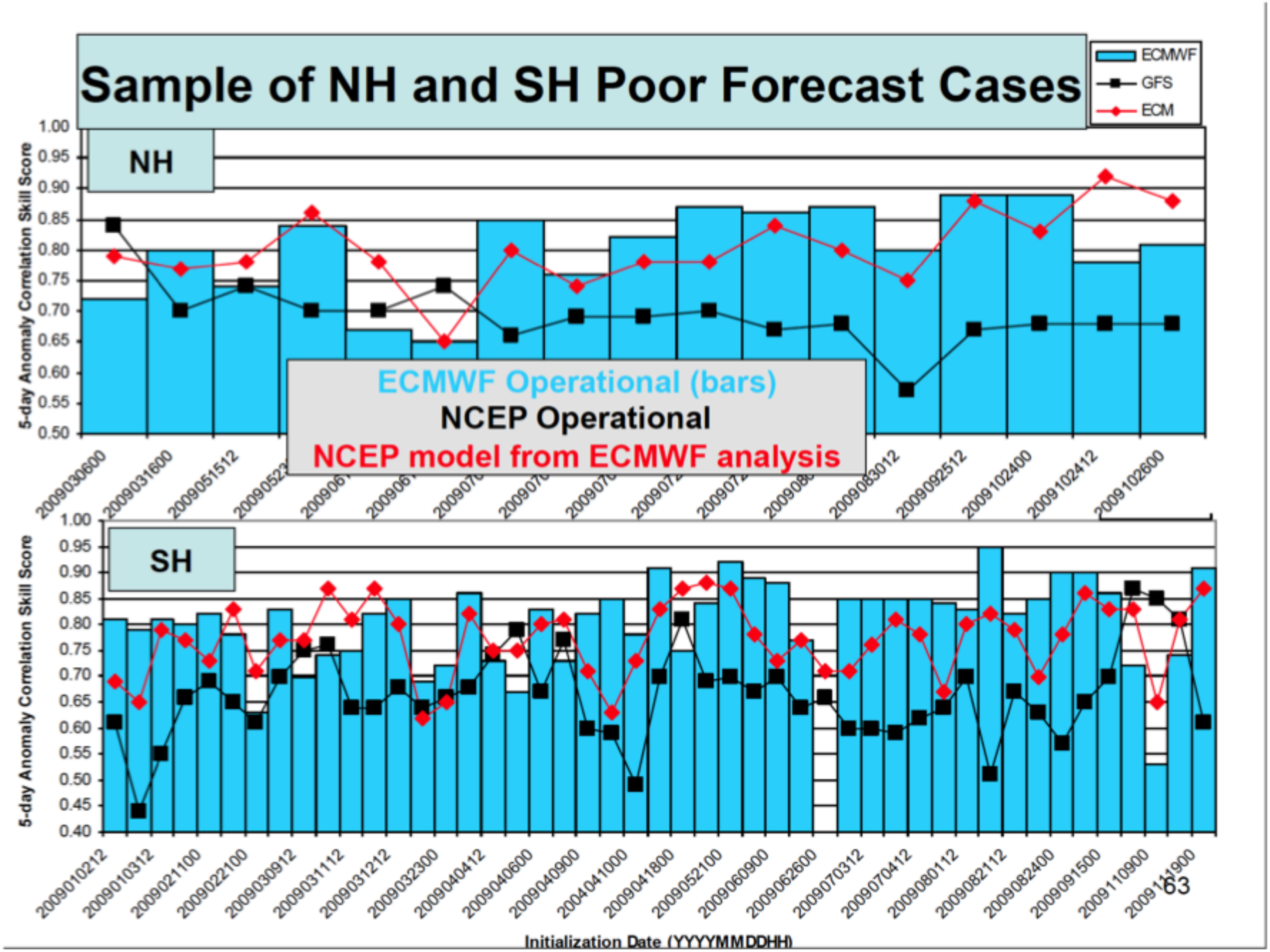
Promising new tools for the LETKF

Forecast Sensitivity to Observations and "proactive QC"

(with Y Ota, T Miyoshi, J Liu, and J Derber)

- This project was started by NCEP findings of the "5-day skill dropouts"
- A simpler, more accurate formulation than Liu and Kalnay (2008) for the Ensemble Forecast Sensitivity to Observations (EFSO, Kalnay et al., 2012, Tellus).
- Ota et al., 2013 tested it with the NCEP EnSRF-GFS operational system using all operational observations.
- Allows to identify "bad observations" after 12 or 24hr, and then repeat the data assimilation without them: "proactive QC".

The NCEP 5-day skill dropout problem

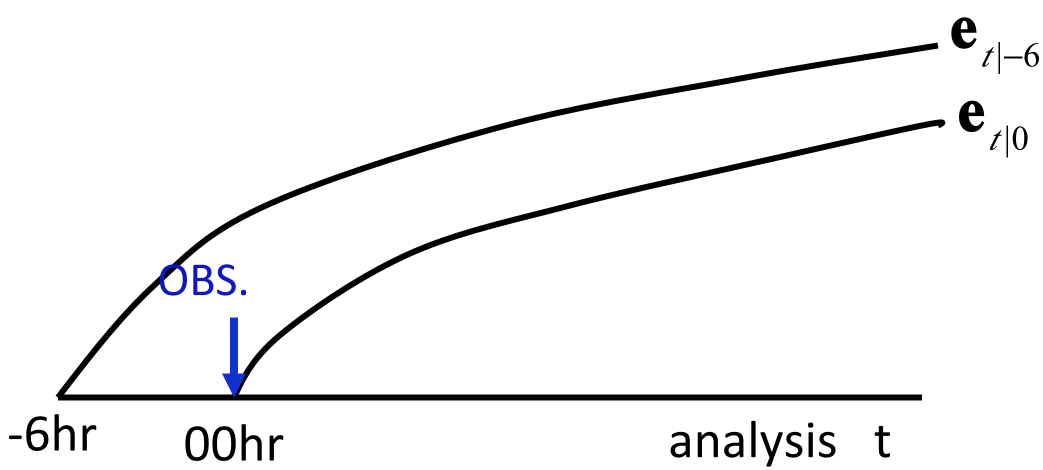


Ensemble Forecast Sensitivity to Observations

"Adjoint sensitivity without adjoint" (Liu and K, 2008, Li et al., 2010)

Here we show a simpler, more accurate formulation

(Kalnay, Ota, Miyoshi: Tellus, 2012)



$$\mathbf{e}_{t|0} = \overline{\mathbf{x}}^f_{t|0} - \overline{\mathbf{x}}^a_t$$

(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:

$$(\overline{\mathbf{x}}^a_0 - \overline{\mathbf{x}}^b_{0|-6}) = \mathbf{K}(\mathbf{y} - H(\mathbf{x}^b_{0|-6}))$$

- Observation impact on the reduction of forecast error:

$$\Delta\mathbf{e}^2 = (\mathbf{e}^T_{t|0}\mathbf{e}_{t|0} - \mathbf{e}^T_{t|-6}\mathbf{e}_{t|-6}) = (\mathbf{e}^T_{t|0} - \mathbf{e}^T_{t|-6})(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

Ensemble Forecast Sensitivity to Observations

$$\Delta \mathbf{e}^2 = (\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T \mathbf{e}_{t|-6}) = (\mathbf{e}_{t|0}^T - \mathbf{e}_{t|-6}^T)(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

$$= (\overline{\mathbf{x}}_{t|0}^f - \overline{\mathbf{x}}_{t|-6}^f)^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

$$= \left[\mathbf{M}(\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_{0|-6}^b)\right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}), \text{ so that}$$

$$\boxed{\Delta \mathbf{e}^2 = \left[\mathbf{MK}(\mathbf{y} - H(\mathbf{x}_{0|-6}^b))\right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})}$$

Langland and Baker (2004), Gelaro, solve this with the adjoint:

$$\boxed{\Delta \mathbf{e}^2 = \left[(\mathbf{y} - H(\mathbf{x}_{0|-6}^b))\right]^T \mathbf{K}^T \mathbf{M}^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})}$$

This requires the adjoint of the model $\mathbf{M}^T$ and of the data assimilation system $\mathbf{K}^T$ (Langland and Baker, 2004)

Ensemble Forecast Sensitivity to Observations

Langland and Baker (2004):

$$\boxed{\begin{aligned}\Delta \mathbf{e}^2 &= \left[\mathbf{MK}(\mathbf{y} - H(\mathbf{x}^b_{0|-6}))\right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\ &= \left[(\mathbf{y} - H(\mathbf{x}^b_{0|-6}))\right]^T \mathbf{K}^T\mathbf{M}^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})\end{aligned}}$$

With EnKF we can use the original equation without "adjointing":

Recall that $\mathbf{K} = \mathbf{P}^a\mathbf{H}^T\mathbf{R}^{-1} = 1/(K-1)\mathbf{X}^a\mathbf{X}^{aT}\mathbf{H}^T\mathbf{R}^{-1}$ so that

$$\mathbf{MK} = \mathbf{MX}^a(\mathbf{X}^{aT}\mathbf{H}^T)\mathbf{R}^{-1}/(K-1) = \mathbf{X}^f_{t|0}\mathbf{Y}^{aT}\mathbf{R}^{-1}/(K-1)$$

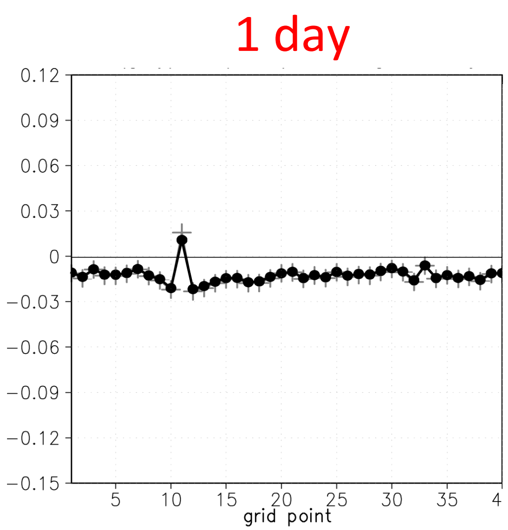
Thus,

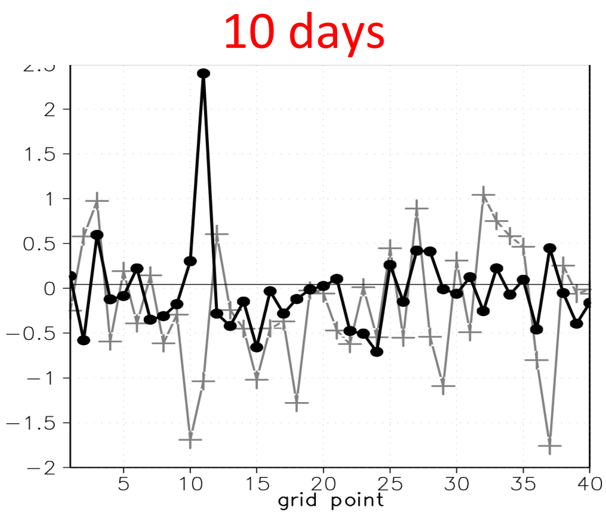
$$\boxed{\begin{aligned}\Delta \mathbf{e}^2 &= \left[\mathbf{MK}(\mathbf{y} - H(\mathbf{x}^b_{0|-6}))\right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\ &= \left[(\mathbf{y} - H(\mathbf{x}^b_{0|-6}))\right]^T \mathbf{R}^{-1}\mathbf{Y}^a_0\mathbf{X}^{fT}_{t|0}(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})/(K-1)\end{aligned}}$$

This uses the available **nonlinear** forecast ensemble products.

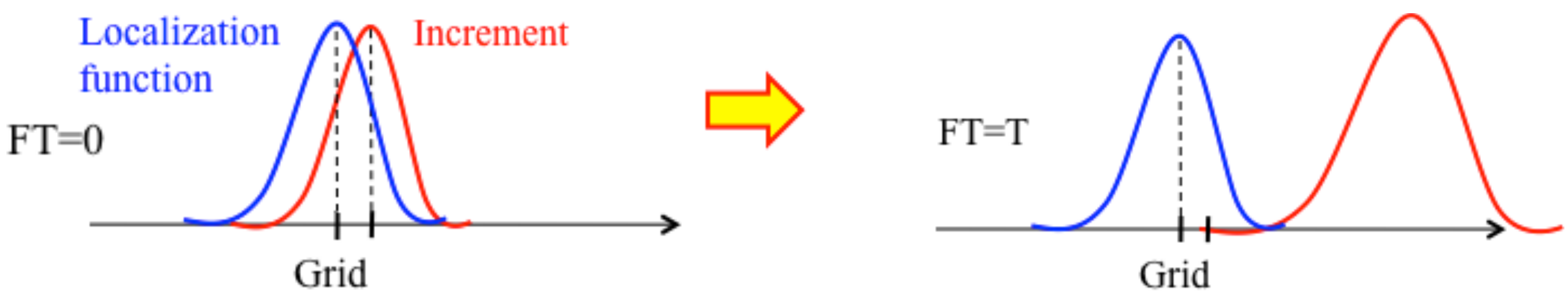
Tested ability to detect a poor quality ob impact on the forecast in the Lorenz 40 variable model

Observation impact from LB(+) and from ensemble sensitivity ( ) ●





- ✓The adjoint and the ensemble sensitivity give similar observation impact on the 24 hr forecast.
- ✓The ensemble sensitivity is nonlinear and is able to detect bad obs for longer forecasts
- ✓This was done ignoring EnKF localization

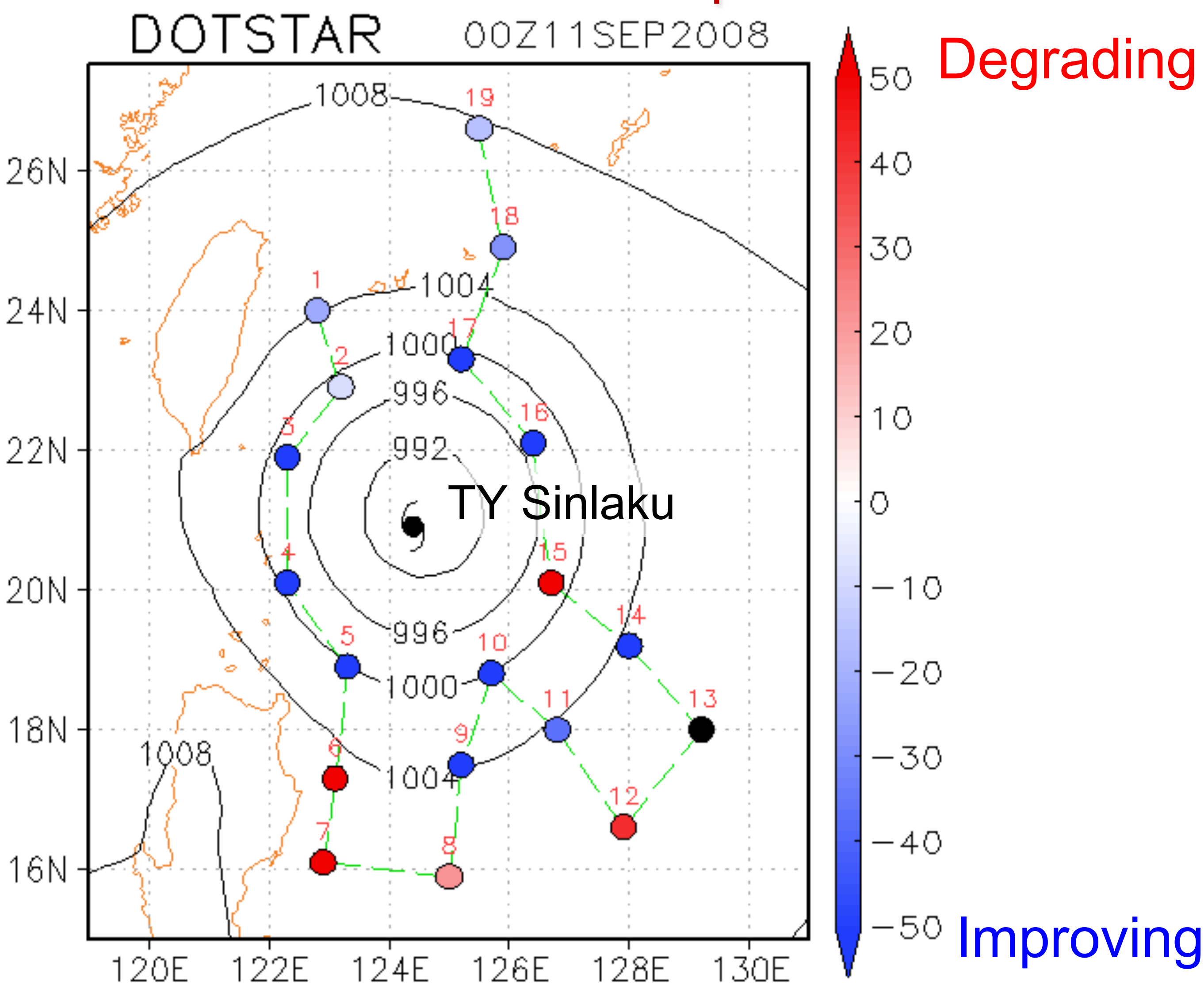


The localization center point for observation impact estimate is now moved with the horizontal wind: an approximation

Impact of dropsondes on a Typhoon

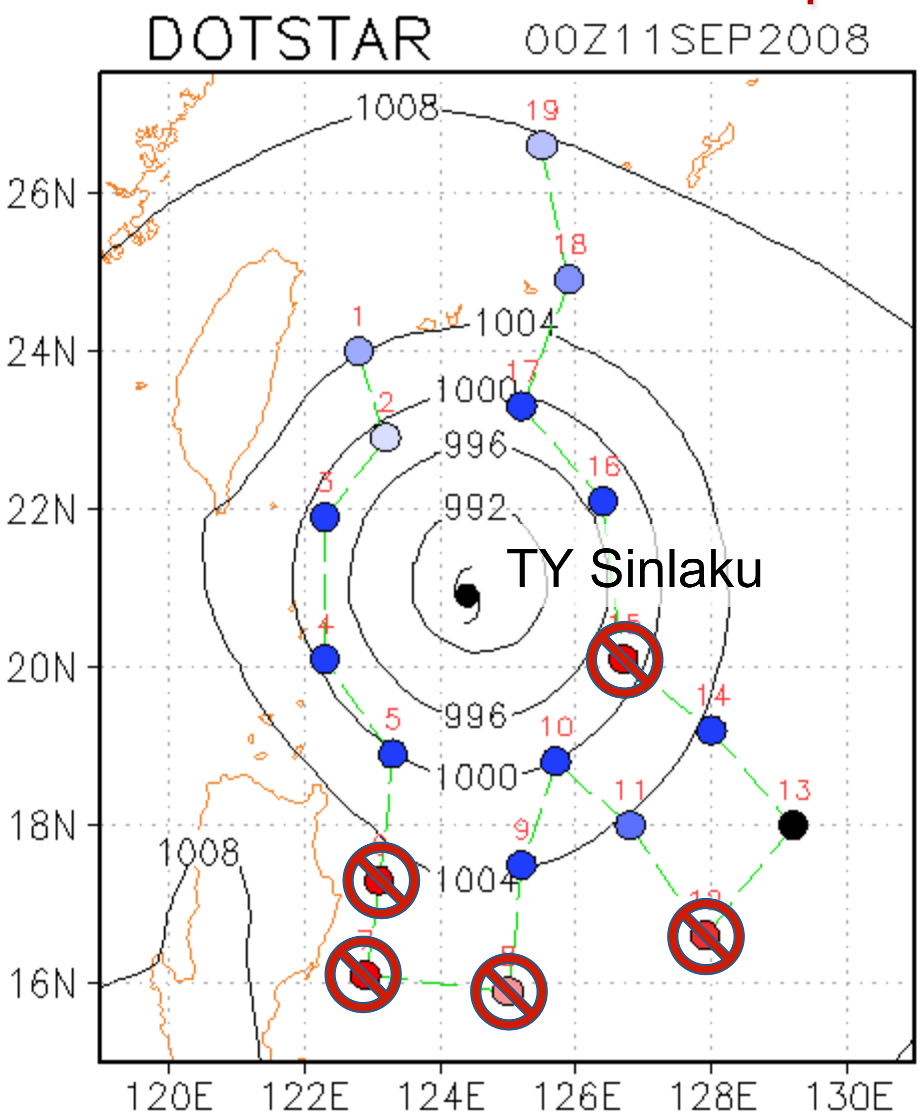
(Kunii et al. 2012)

Estimated observation impact

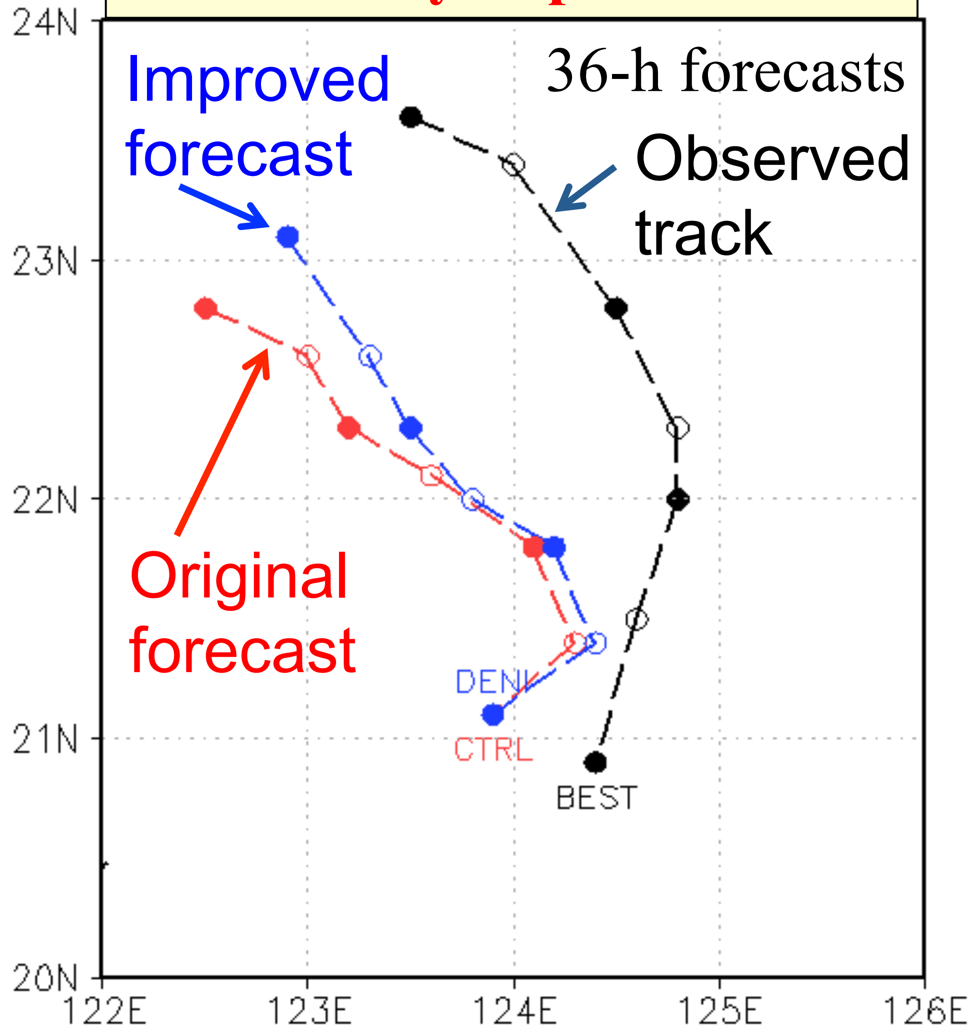


Denying negative impact data improves forecast!

Estimated observation impact



Typhoon track forecast is actually improved!!



Ota et al. 2013: Applied EFSO to NCEP GFS/EnSRF using all operational observations. Determined regional 24hr "forecast failures"

- Divide the globe into 30x30^{o} regions

- Find all cases where the 24hr regional forecast error is at least 20% larger than the 36hr forecast error verifying at the same time, and

- where the 24hr forecast has errors at least twice the time average.

- Identify the top observation type that has a negative impact on the forecast.

- Found 7 cases of 24hr forecast

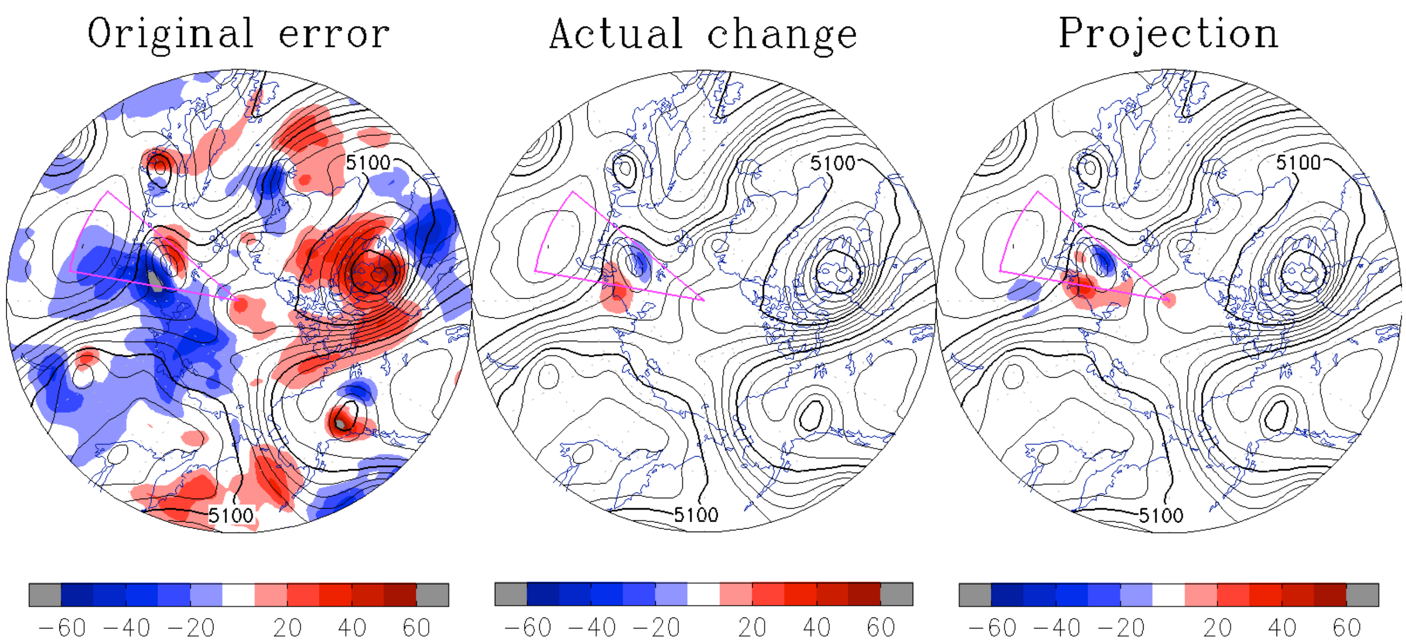
24-hr forecast error correction (Ota et al.)

- identified 7 cases of large 30°x30° regional errors,
- rerun the forecasts denying bad obs.
- the forecast errors were substantially reduced
- this could be applied to improve the 5-day skill dropouts

MODIS

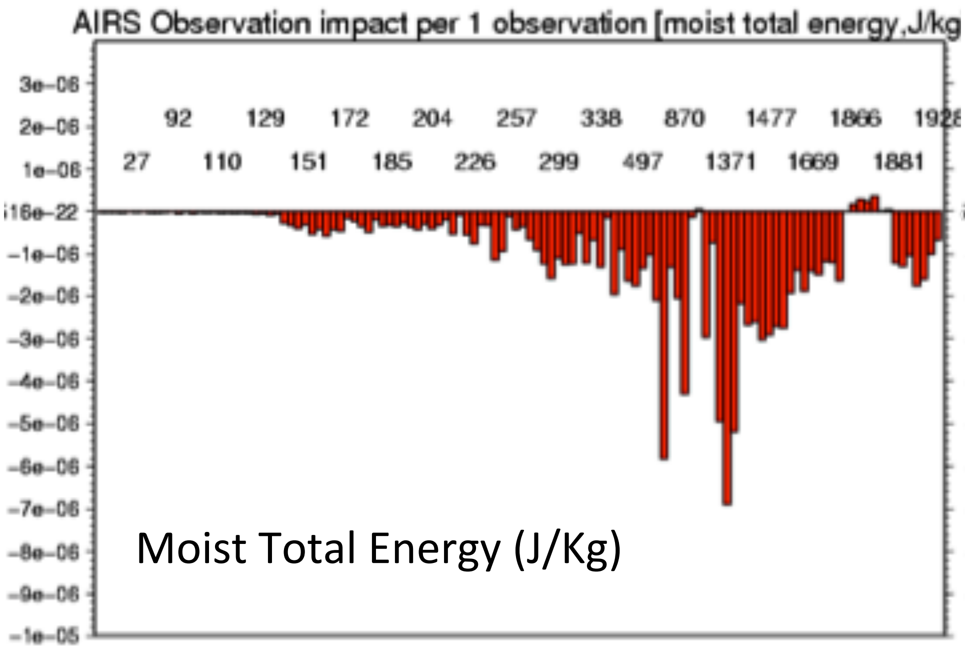
Initial	Area	Size	Rate	N	Denied observation	Change
12 UTC JAN 10	90S~60S 100E~130E	2.04	1.20	1	GPSRO (80S~60S, 90E~120E) ASCAT (60S~50S, 100E~120E)	-6.6%
06 UTC JAN 12	50N~80N 150E ~ 180	2.18	1.40	1	AMSUA (ch4: 45N~75N, 160E~170W, ch5:40N~55N, 155E~180, NOAA15 ch6: 50N~75N, 140E~170W, ch7: 70N~80N, 130E~170E)	-11.4%
00 UTC JAN 16	30N~60N 30W~0	2.13	1.31	2	Radiosonde wind (Valentia, Ireland), ASCAT (40N~47N, 20W~10W, 50N~55N, 35W~30W)	-1.0%
12 UTC JAN 22	90S~60S 130E~160E	2.34	1.22	2	AMSUA (ch5: 65S~50S, 90E~110E, 60S~50S, 120E~127E, ch6: 60S~45S, 110E~125E)	-2.2%
06 UTC FEB 2	50N~80N 150W~120W	3.10	1.32	4	IASI (35N~45N, 155W~150W) NEXRAD (55N~60N, 160W~135W)	-5.5%
18 UTC FEB 6	60N~90N 50E~80E	2.06	1.71	2	MODIS_Wind (60N~90N, 30E~90E)	-39.0%
18 UTC FEB 6	90S~60S 20W~10E	3.56	1.22	1	MODIS_Wind (80S~50S, 30W~0)	-22.5%

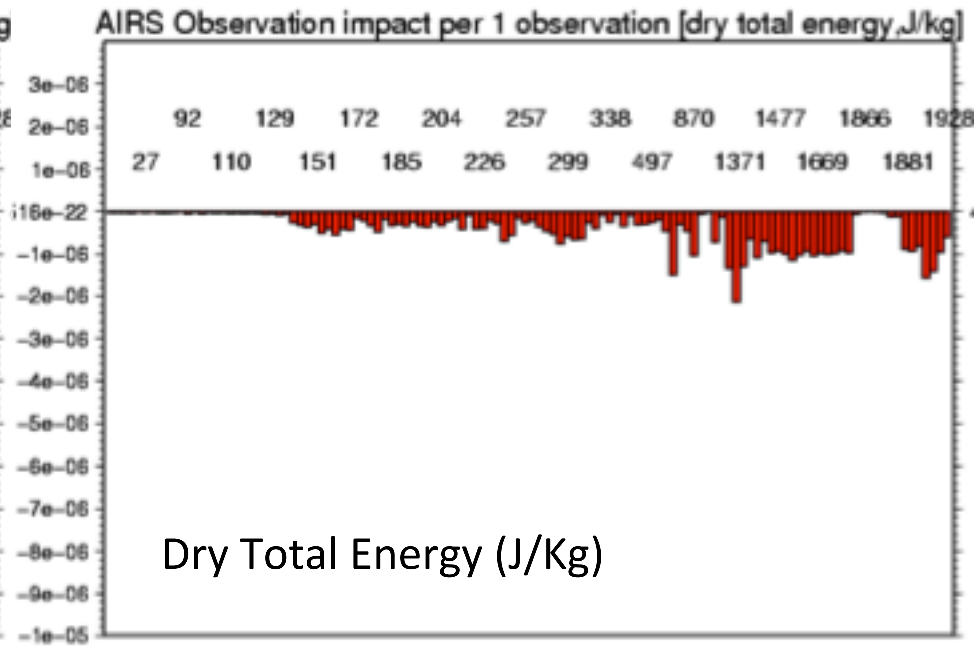
“Proactive” QC: Bad observations can be identified by EFSO and withdrawn from the data assimilation



After identifying MODIS polar winds producing bad 24 hr regional forecasts, the withdrawal of these winds reduced the forecast errors by 39%, as projected by EFSO.

Other applications: Impacts of Observing Systems





The EnKF formulation is nonlinear and thus allows computing Moist Total Energy and estimate more accurately the impact of the channels on the moisture forecast. Adjoint formulation needs TLM.

Summary

- The new EFSO formulation works well and uses available EnKF products.
- It can be used to detect observations that give bad regional 12hr or 24hr forecasts.
- We can then repeat the data assimilation without the bad obs, a powerful tool for a “proactive” QC and monitoring.
- Operationally, it should be possible to accumulate flawed observations with detailed information about the atmospheric characteristics and how the observations were used.
- This should give enough information to the observation model developers to improve the algorithms and avoid similar problems.

Promising new tools for the LETKF

Application of ensemble forecast sensitivity to data assimilation

(Shu-Chih Yang, E. Kalnay, with thanks to T. Enomoto)

Ensemble Sensitivity: Application to Data Assimilation and the Spin-up Problem

Assume we are in a window of the LETKF with an ensemble of K members

$$\mathbf{x}_{i,t}^{b} = M(\mathbf{x}_{i,t-1}^{a})$$

Since the window is short,

$$\delta\mathbf{x}_{i,t}^{b} = \mathbf{x}_{i,t}^{b} - \overline{\mathbf{x}}_{t}^{b} \approx \mathbf{M}(\delta\mathbf{x}_{i,t-1}^{a})$$

Define the vectors of analysis and forecast perturbations:

$$\mathbf{X}_{t-1}^{a} = [\delta\mathbf{x}_{1,t-1}^{a}, \ldots, \delta\mathbf{x}_{K,t-1}^{a}]; \quad \mathbf{X}_{t}^{b} = [\delta\mathbf{x}_{1,t}^{b}, \ldots, \delta\mathbf{x}_{K,t}^{b}]$$

We want to find the linear combination of analysis perturbations that will grow fastest (Singular Vectors):

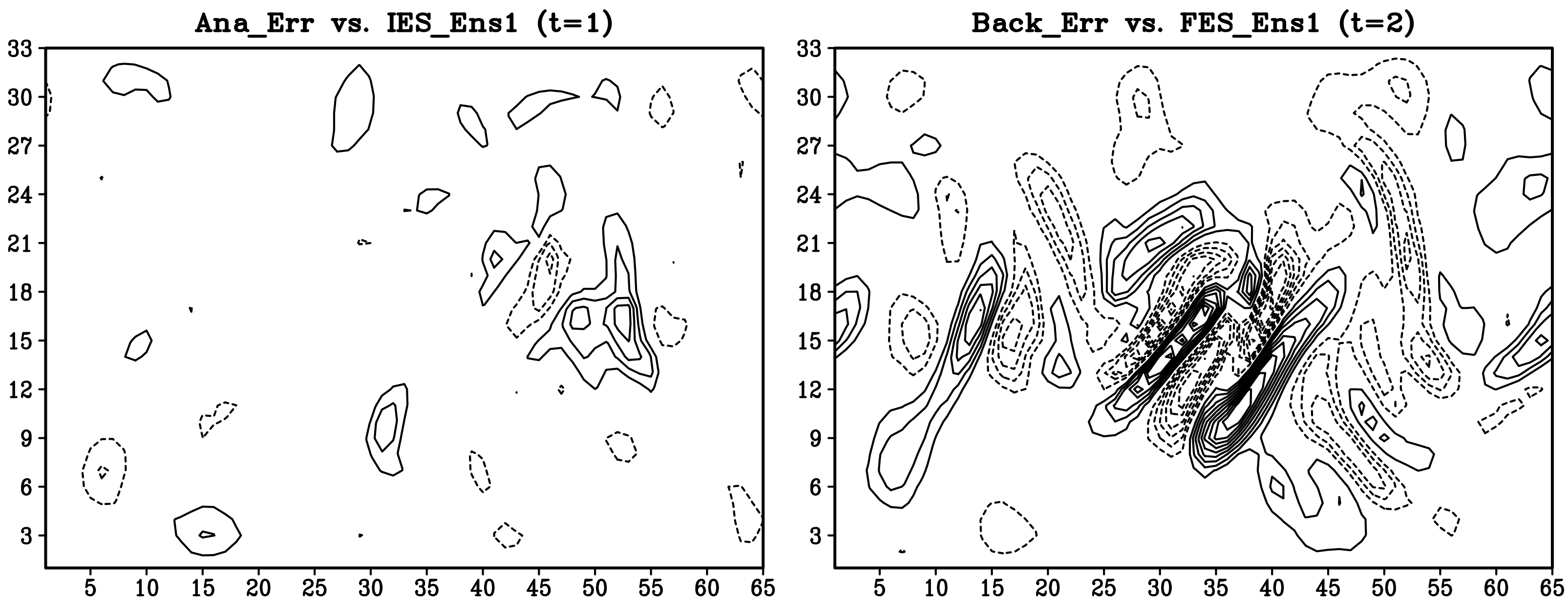
$$\delta \mathbf{x}_{t-1}^{a} = \mathbf{X}_{t-1}^{a}\mathbf{p}; \qquad \delta \mathbf{x}_{t}^{b} = \mathbf{X}_{t}^{b}\mathbf{p}$$

with optimal coefficients $\mathbf{p} = [\, p_{t,1}, \ldots., p_{t,K} \,]$

We can use the equation in Enomoto et al (2007) (see derivation in Yang and Kalnay, 2013):

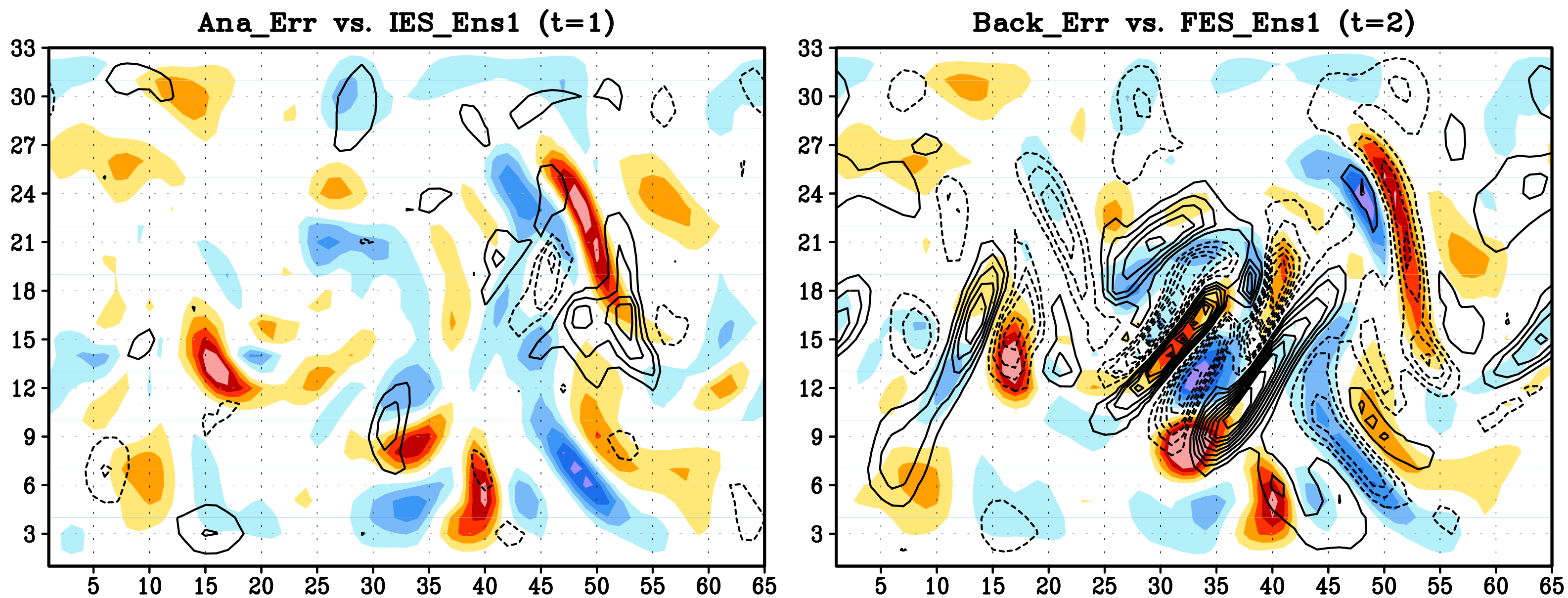
$$(\mathbf{X}_{t-1}^{aT} C_I \mathbf{X}_{t-1}^{aT})^{-1} (\mathbf{X}_{t}^{bT} C_F \mathbf{X}_{t}^{bT})\mathbf{p} = \lambda \mathbf{p}$$

We tested this with a QG model starting with a random ensemble that satisfies the $B_{3D\text{-}Var}$.
The initial optimal perturbation after only 6hr grows into a final perturbation after 12 hrs:



Is this fast growing perturbation related to the background errors?

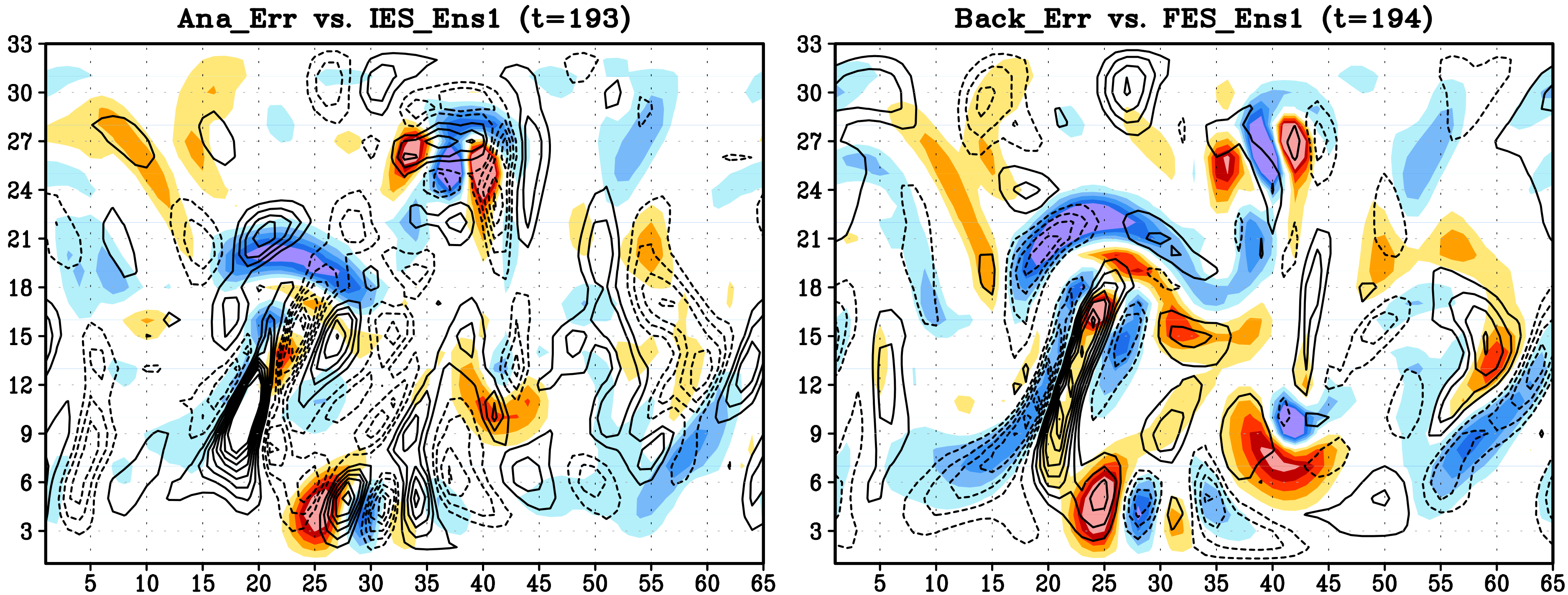
We tested this with a QG model starting with a random ensemble that satisfies the $B_{3D\text{-}Var}$.
The initial optimal perturbation after 6hr grows into a final perturbation after 12 hrs:



Is this fast growing perturbation related to the background errors?

YES!!!

Later in the run, the relationship between the final SVs and the forecast errors is even stronger



We can use the method of Kalnay and Toth (1994):
"Removing growing errors in the analysis cycle"

We improved the first guess by finding μ such that

$$\mathbf{O} - (\mathbf{F} + \mu\mathbf{G}) \perp \mathbf{G}$$

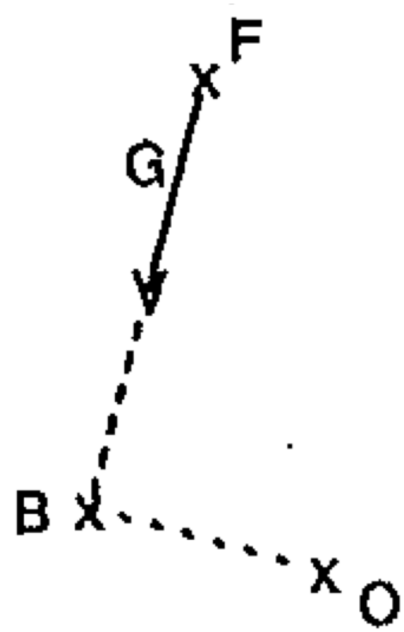


Fig. 1: Schematic for the adjustment of the first guess F towards the observations O in the direction of the growing errors G.

We used one BV as the growing mode G, and found μ

locally every 10X10 degrees, and interpolated in between.

The results showed a remarkable improvement in the forecasts!

We could use the SVs as the growing modes.

9th International Carbon Dioxide Conference
Jun. 3-7, 2013

KIAPS
www.kiaps.org

Estimation of Surface CO_2 Fluxes from Atmospheric Data Assimilation

[1]Ji-Sun Kang, [2]Eugenia Kalnay, [3]Junjie Liu, [4]Inez Fung, and [5]Takemasa Miyoshi

[1]Korea Institute of Atmospheric Prediction Systems, Seoul, Korea
[2]University of Maryland, College Park, MD, USA
[3]NASA/JPL, Pasadena, CA, USA
[4]University of Califonia, Berkeley, CA, USA
[5]Advanced Institute of Computational Science, RIKEN, Kobe, Japan

Korea Institute of Atmospheric Prediction Systems (KIAPS)
(재)한국형수치예보모델개발사업단

LETKF-C with SPEEDY-C

- Model: **SPEEDY-C** (Molteni, 2003; Kang, 2009)
  - Spectral AGCM model with T30L7
  - Prognostic variables: U, V, T, q, Ps, C
    - ✓ C (atmospheric CO_2): an inert tracer
  - Persistence forecast of Carbon Fluxes (CF), no observations
- Simulated observations
  - Rawinsonde observations of U, V, T, q, Ps
  - Ground-based observations of atmospheric CO_2
    - ✓ 18 hourly and 107 weekly data on the globe
  - Remote sensing data of column mixing CO_2
    - ✓ **AIRS** whose averaging kernel peaks at mid-troposphere
    - ✓ **GOSAT** whose averaging kernel is nearly uniform throughout the column
- Initial condition: random (no *a-priori* information)
- 20 ensembles

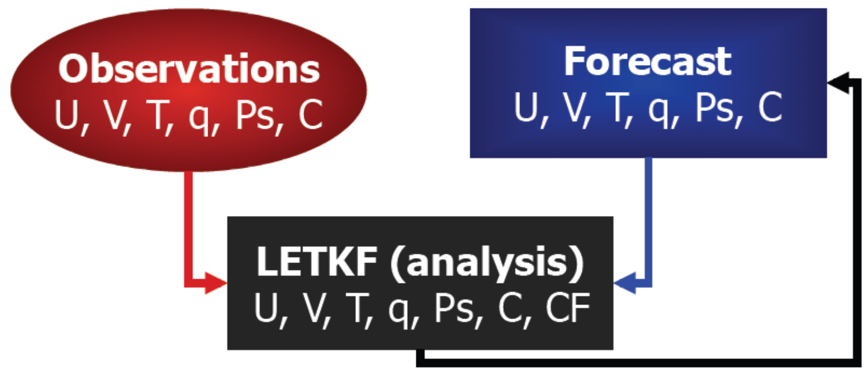
Surface flux estimation within EnKF

- Parameter estimation using state vector augmentation

$$\mathbf{X}^b = \begin{bmatrix} \mathbf{X} \\ \mathbf{CF} \end{bmatrix}$$

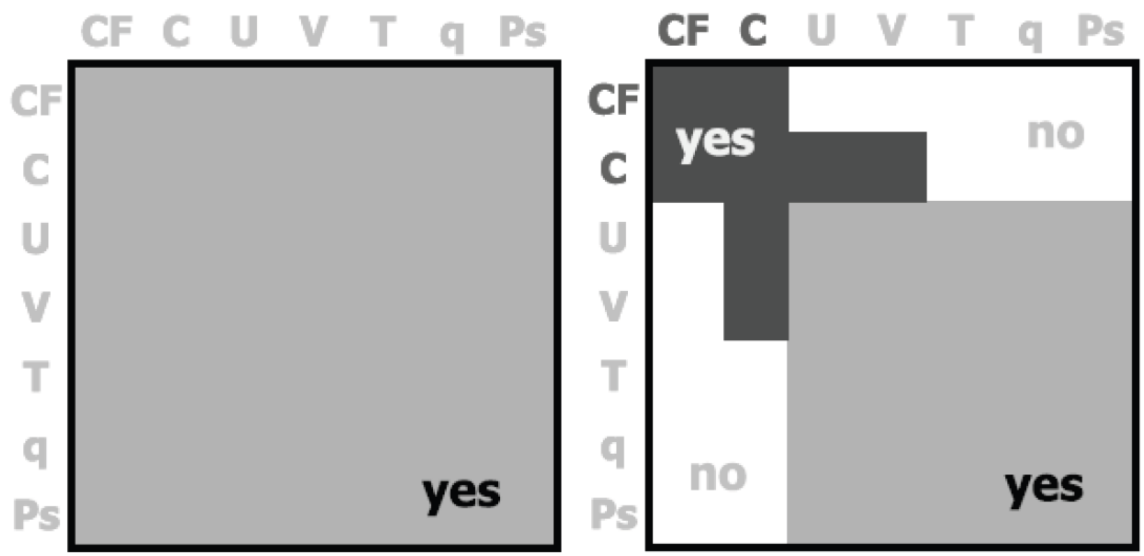
: model state vector (U, V, T, q, Ps, C)

: surface CO_2 flux

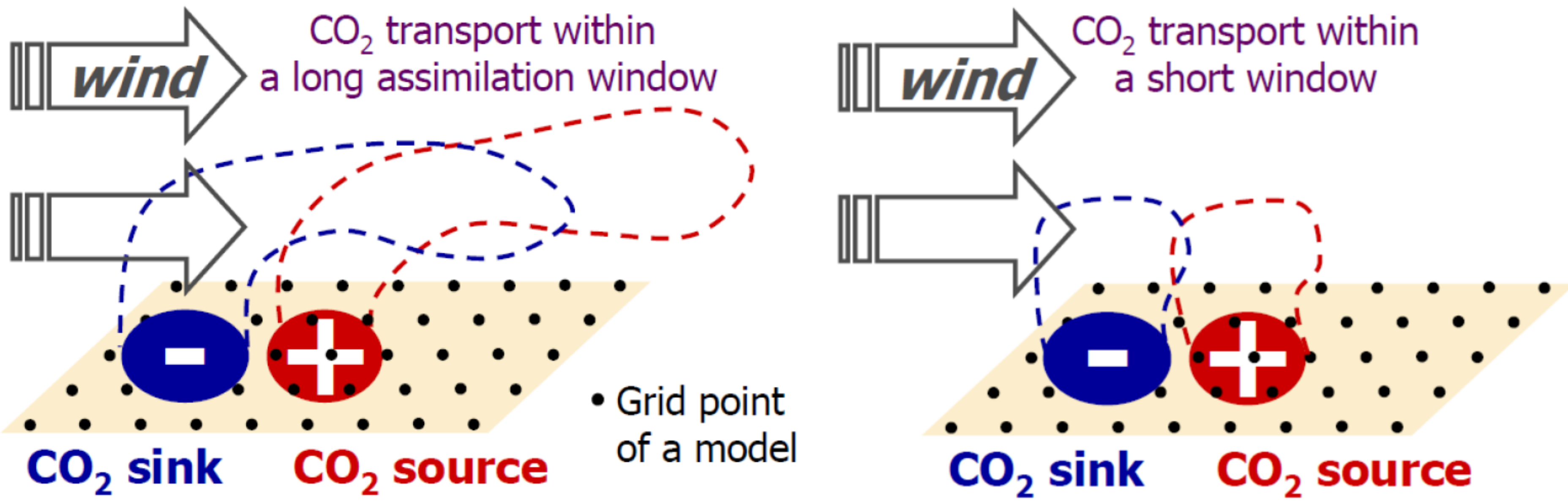


- Append CF
- Update CF as a part of the data assimilation processes
- Multivariate analysis with a localization of the variables (Kang et al. 2011, JGR)

Schematic plots of background error covariance matrix $\mathbf{P}^b$ → without *"variable localization"* (left) and with it (right)



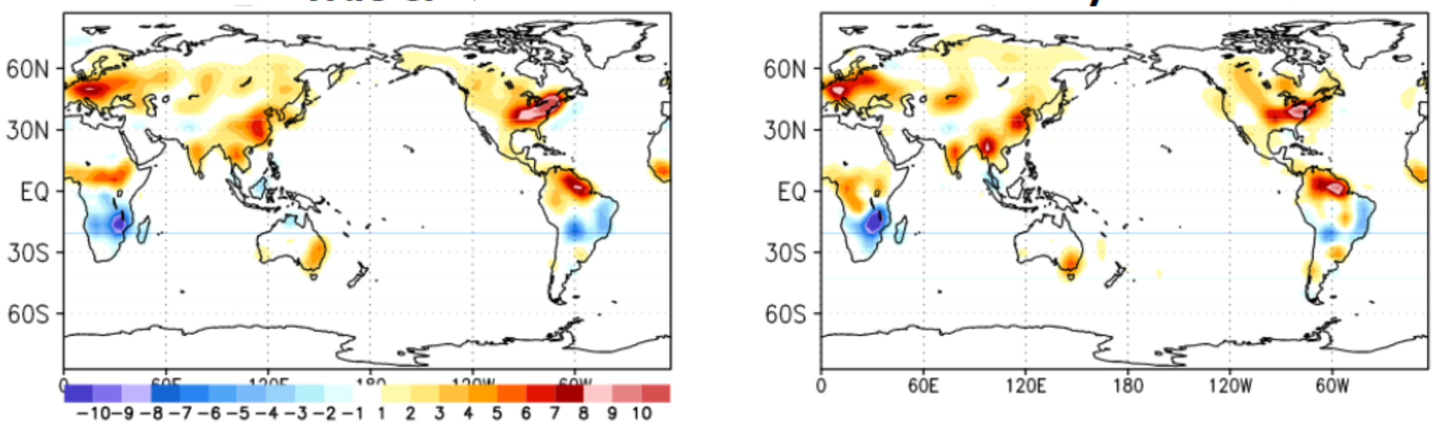
Assimilation window in LETKF-C



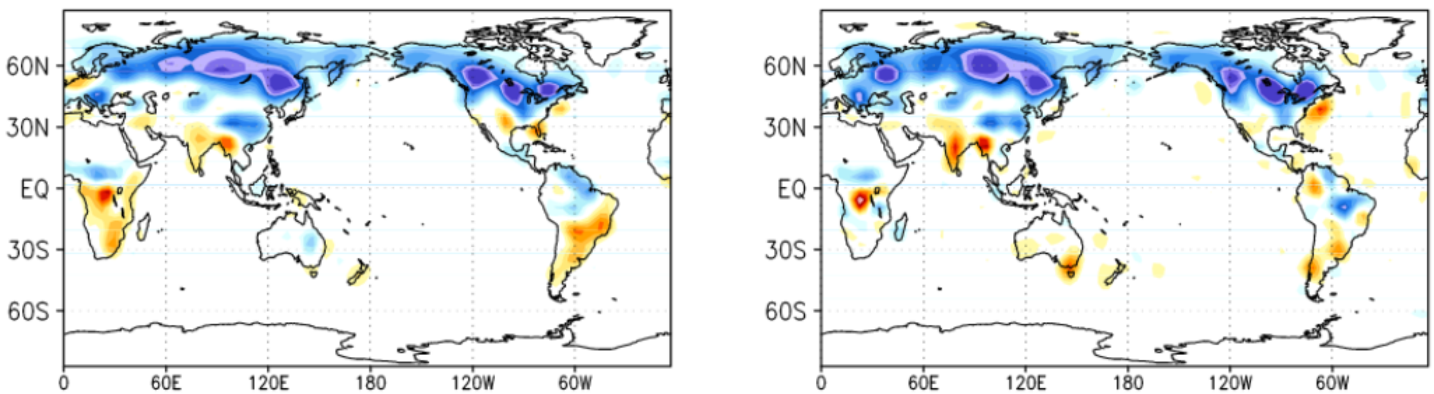
- A short assimilation window reduces the attenuation of observed CO_2 information because the analysis system can use the strong correlation between C and CF before the transport of C blurs out the essential information of CF forcing
- We may not be able to reflect the optimal correlation between C and CF within a long assimilation window, which can introduce sampling errors into the EnKF analysis

Results

▼ True CF ▼ ▼ Analysis of CF ▼

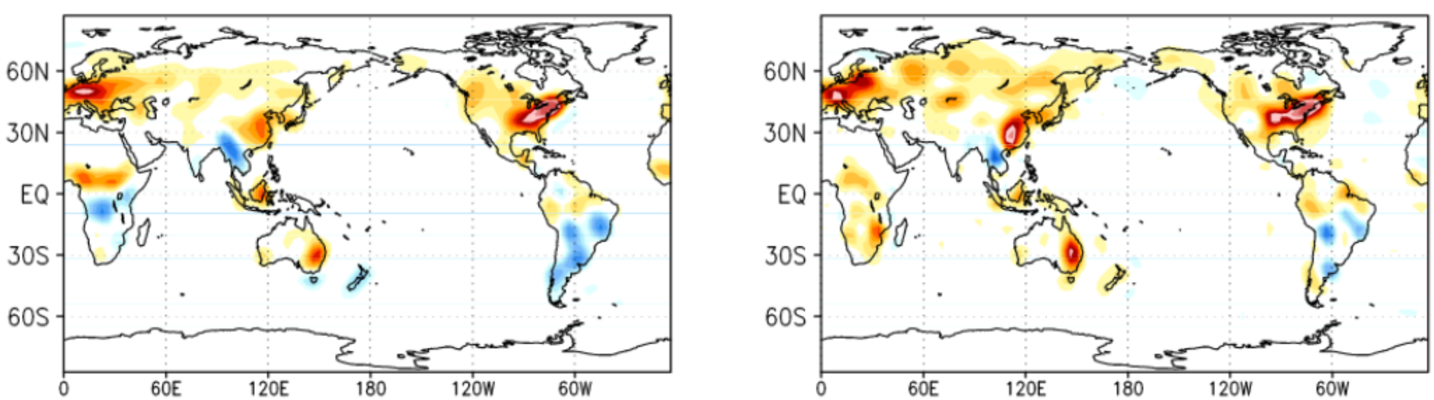


◀ 00Z01APR
after 3 months of DA



◀ 00Z01AUG
after 7 months of DA

succeeded in estimating time-evolving CF at model-grid scale



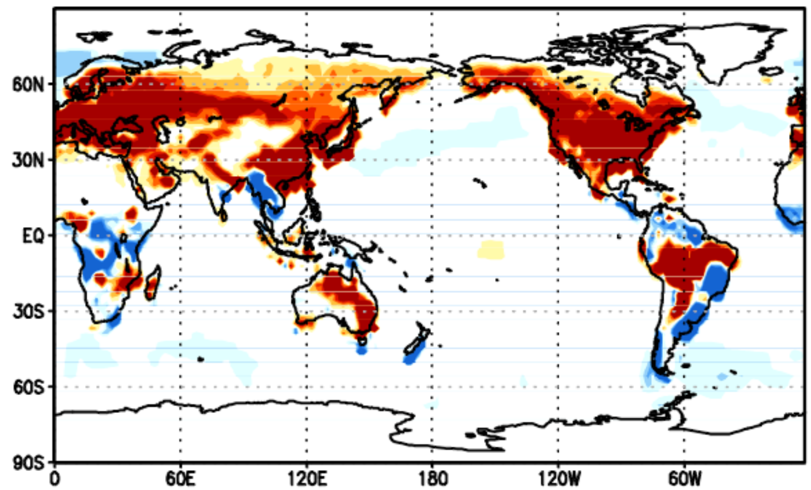
◀ 00Z01JAN
after one year of DA

LETKF-C with NCAR CAM3.5

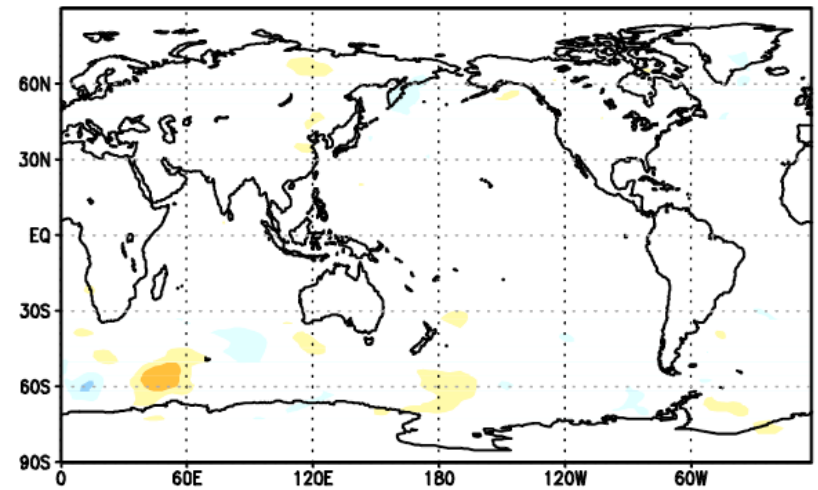
- Model: **CAM 3.5**
  - Finite Volume dynamical core
  - 2.5°×1.9° of horizontal resolution with 26 layers in the vertical
  - C (atmospheric CO_2) is an inert tracer
  - Persistence forecast of CF
- Simulated observations with **real observation coverage**
  - Conventional data for U, V, T, q, Ps
  - Ground–based observations of atmospheric CO_2
    - ~10 hourly and ~100 weekly records on the globe
  - Remote sensing data of column mixing CO_2
    - **AIRS** whose averaging kernel peaks at mid-troposphere
- Initial conditions: random (no *a-priori* information)
- 64 ensembles

LETKF-CAM3.5 CF analysis

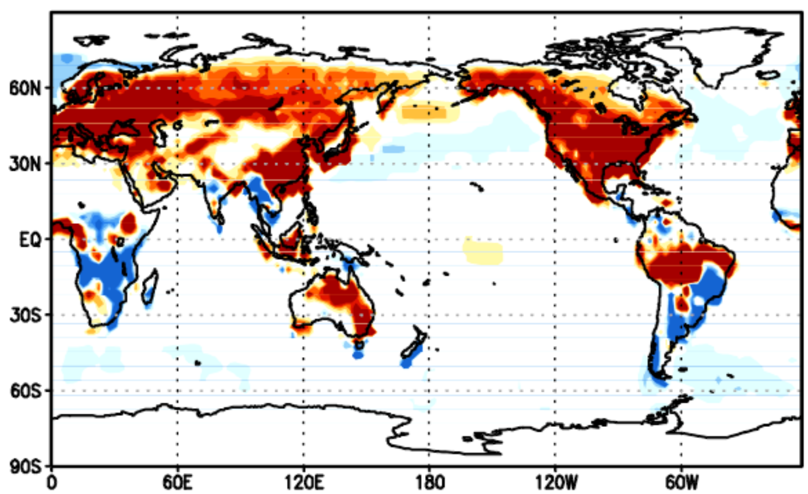
True CF @ initial time (00Z01JAN)



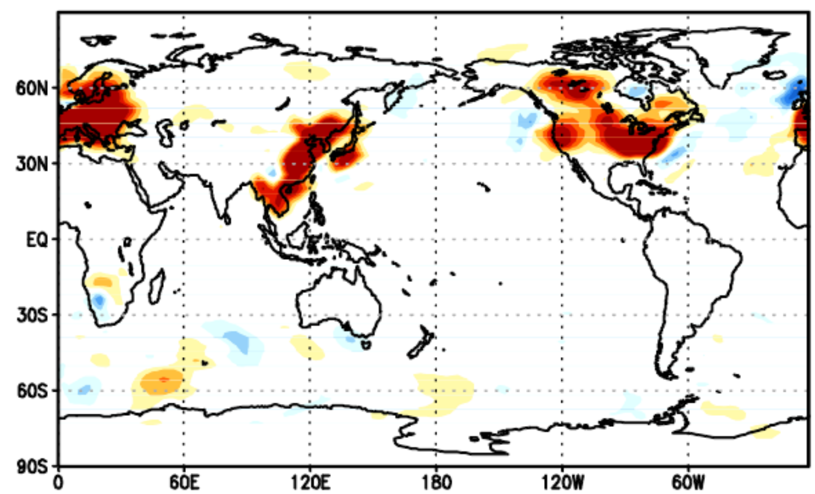
Initial CF



True CF @ 00Z27JAN)



CF analysis @ 00Z27JAN

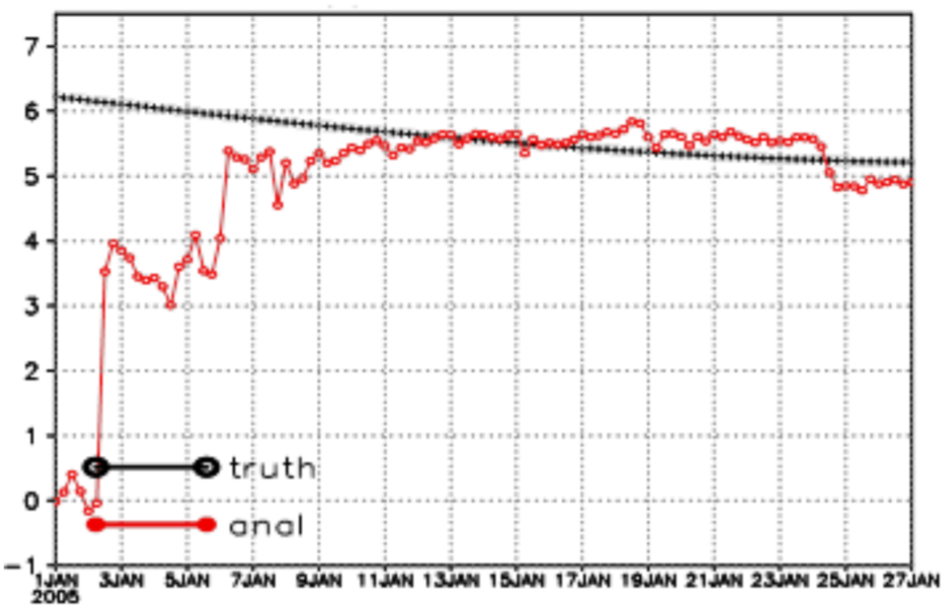


KIAPS

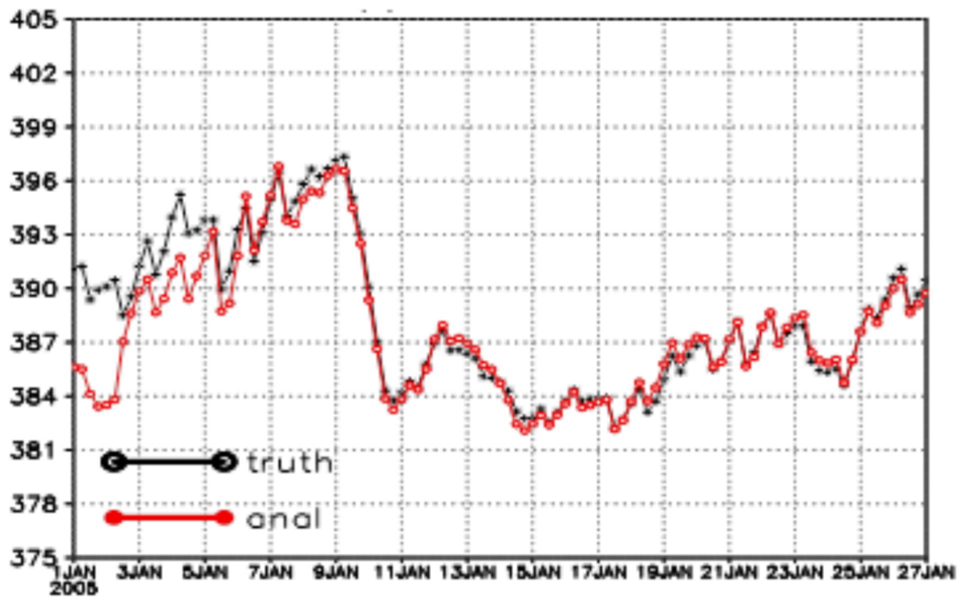
LETKF-CAM3.5 CF analysis

- Time series of surface CO_2 fluxes and atmospheric CO_2 concentration over Europe (observation-rich area)

(a) Surface CO_2 fluxes over EUR



(b) Atmospheric CO_2 over EUR



Summary of LETKF-C carbon fluxes

- We succeeded in estimating surface CO_2 fluxes with the advanced simultaneous analysis system of LETKF-C, even without a-priori information (OSSEs with SPEEDY model)
  - Localization of the variables (Kang et al., 2011, JGR)
  - Advanced data assimilation techniques such as adaptive multiplicative and additive inflation, vertical localization of column mixing CO_2 data (Kang et al., 2012, JGR)
  - EnKF has better performance with a short window
    - CO_2 observations may be able to provide some information to distant CF, but it becomes blurred.
- On-going work of LETKF-C with CAM3.5
  - OSSEs with real observation coverages has been examined
  - Preliminary results are encouraging.
- The same methodology has been applied to estimating surface heat, moisture, and momentum fluxes
  - Results are promising. (Kang et al., 2013, in prep)

Summary of LETKF-C carbon fluxes

- We succeeded in estimating surface CO_2 fluxes with the advanced simultaneous analysis system of LETKF-C, even without a-priori information (OSSEs with SPEEDY model)
  - Localization of the variables (Kang et al., 2011, JGR)
  - Advanced data assimilation techniques such as adaptive multiplicative and additive inflation, vertical localization of column mixing CO_2 data (Kang et al., 2012, JGR)
  - EnKF has better performance with a short window
    - CO_2 observations may be able to provide some information to distant CF, but it becomes blurred.
- On-going work of LETKF-C with CAM3.5
  - OSSEs with real observation coverages has been examined
  - Preliminary results are encouraging.
- The same methodology has been applied to estimating surface heat, moisture, and momentum fluxes
  - Results are promising. (Kang et al., 2013, in prep)

THANK YOU VERY MUCH FOR YOUR ATTETION!

EnKF is a newer, much simpler technology.

There is much more potential not yet exploited or not even explored such as:

- **Estimation and correction of model errors and parameters**
- **Estimation of observation errors**
- **Reducing growing errors from the initial conditions**
- **Accelerating spin-up**
- **…**