Recent advances in EnKF

Eugenia Kalnay, Shu-Chih Yang, Steve Penny, Guo-Yuan Lien, Yoichiro Ota, Ji-Sun Kang, Takemasa Miyoshi, Junjie Liu, Kayo Ide, Brian Hunt, and other collaborators at the University of Maryland **Weather & Chaos** group

Recent advances in EnKF

Eugenia Kalnay, Shu-Chih Yang, Steve Penny, Guo-Yuan Lien, Yoichiro Ota, Ji-Sun Kang, Takemasa Miyoshi, Junjie Liu, Kayo Ide, Brian Hunt, and other collaborators at the University of Maryland **Weather & Chaos** group

Data Assimilation Faculty Search at UMD!!

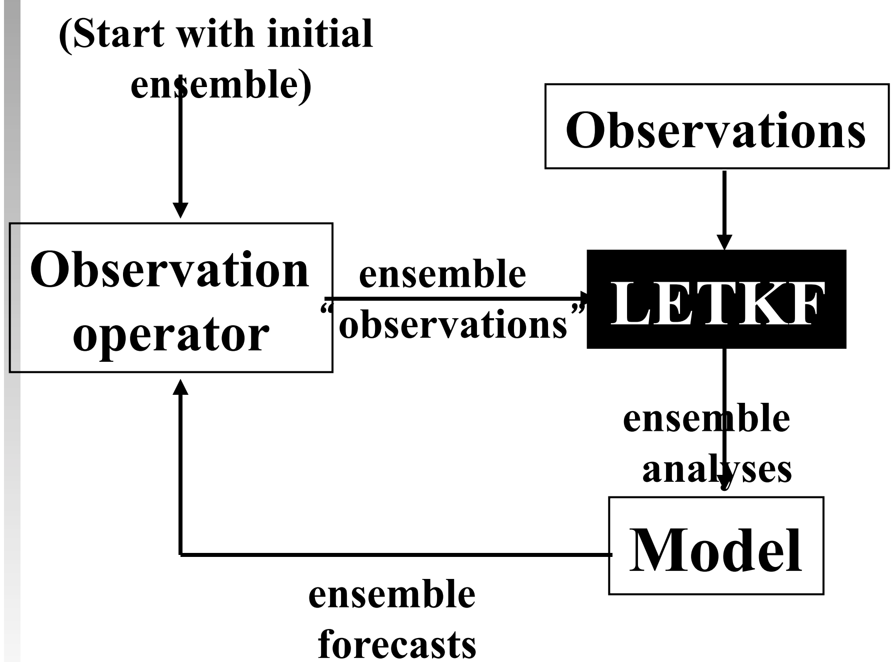
Outline

Review of a few **recent advances** in LETKF

- Running in Place
- Effective assimilation of precipitation
- Ensemble Forecast Sensitivity to Observations (EFSO
- Parameter estimation and carbon cycle data assimilation

- Estimation of **surface heat and moisture fluxes**
- Sensible and latent heat fluxes (SHF, LHF)
- Estimation of **wind stress** in addition to SHF and LHF
- Future Plans

4D-Local Ensemble Transform Kalman Filter (Ott et al, 2004, Hunt et al, 2004, 2007)

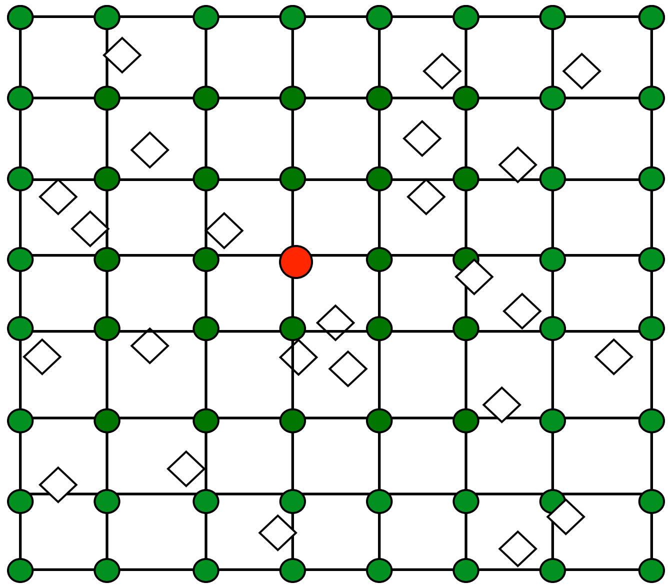


- Model independent (black box)
- Obs. assimilated simultaneously at each grid point
- 100% parallel
- No adjoint needed
- 4D LETKF extension
- Computes the weights for the ensemble forecasts explicitly

Localization based on observations

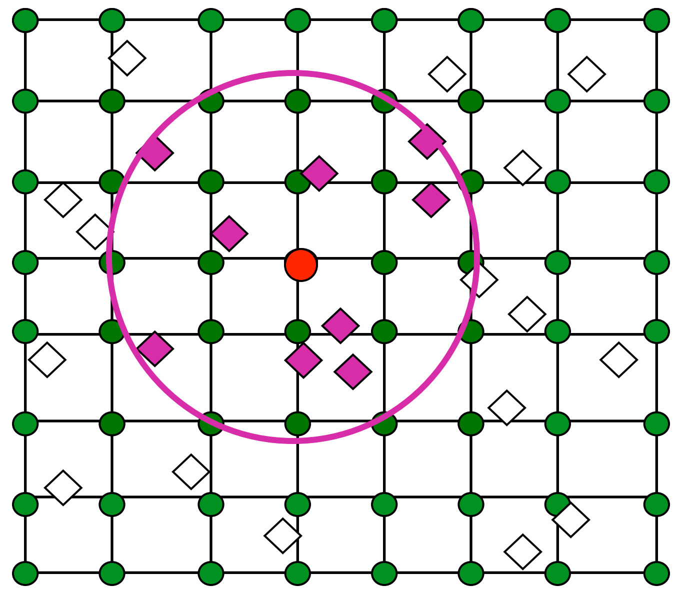
Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot



Localization based on observations

Perform data assimilation in a local volume, choosing observations



The state estimate is updated at the central grid red dot

All observations (purple diamonds) within the local region are assimilated

The LETKF algorithm can be described in a single slide!

Local Ensemble Transform Kalman Filter (LETKF)

Globally:

Forecast step: $\mathbf{x}^b_{n,k} = M_n\left(\mathbf{x}^a_{n-1,k}\right)$

Analysis step: construct $\mathbf{X}^b = \left[\mathbf{x}^b_1 - \overline{\mathbf{x}}^b \mid \ldots \mid \mathbf{x}^b_K - \overline{\mathbf{x}}^b\right];$

$$\mathbf{y}^b_i = H(\mathbf{x}^b_i);\ \mathbf{Y}^b_n = \left[\mathbf{y}^b_1 - \overline{\mathbf{y}}^b \mid \ldots \mid \mathbf{y}^b_K - \overline{\mathbf{y}}^b\right]$$

Locally: **Choose for each grid point the observations to be used, and compute the local analysis error covariance and perturbations in ensemble space:**

$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\right]^{-1};\ \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

Analysis mean in ensemble space: $\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$

and add to $\mathbf{W}^a$ **to get the analysis ensemble in ensemble space.**

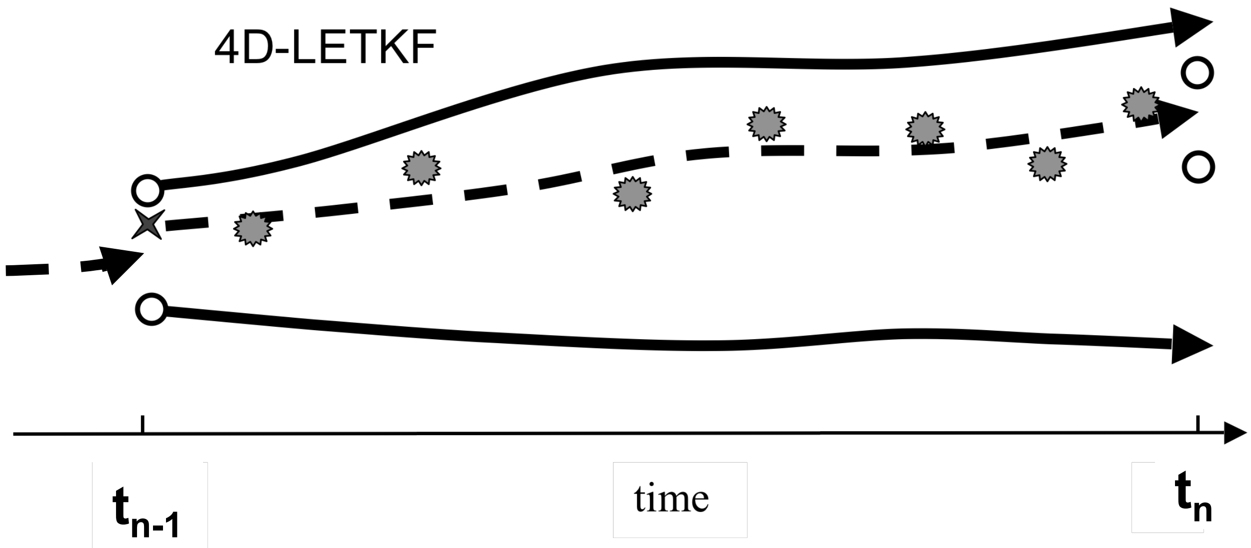
The new ensemble analyses in model space are the columns of $\mathbf{X}^a_n = \mathbf{X}^b_n\mathbf{W}^a + \overline{\mathbf{x}}^b$ **. Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights** $\overline{\mathbf{w}}^a$ **and perturbation analysis matrices of weights** $\mathbf{W}^a$ **. These weights multiply the ensemble forecasts.**

Promising new tools for the LETKF (1)

1. Running in Place (Kalnay and Yang, QJ 2010, Yang, Kalnay, and Hunt, MWR, 2012)

- It extracts more information from observations by using them more than once.
- Useful during spin-up (e.g., hurricanes and tornados).
- It uses the "no-cost smoother", Kalnay et al., Tellus, 2007b.
- Typhoon Sinlaku (Yang et al., 2012, 2013)
- 7-years of Ocean Reanalysis (Penny, 2011, Penny et al., 2013)

No-cost LETKF smoother (×): apply at t_{n-1} the same weights found optimal at t_n. It works for 3D- or 4D-LETKF



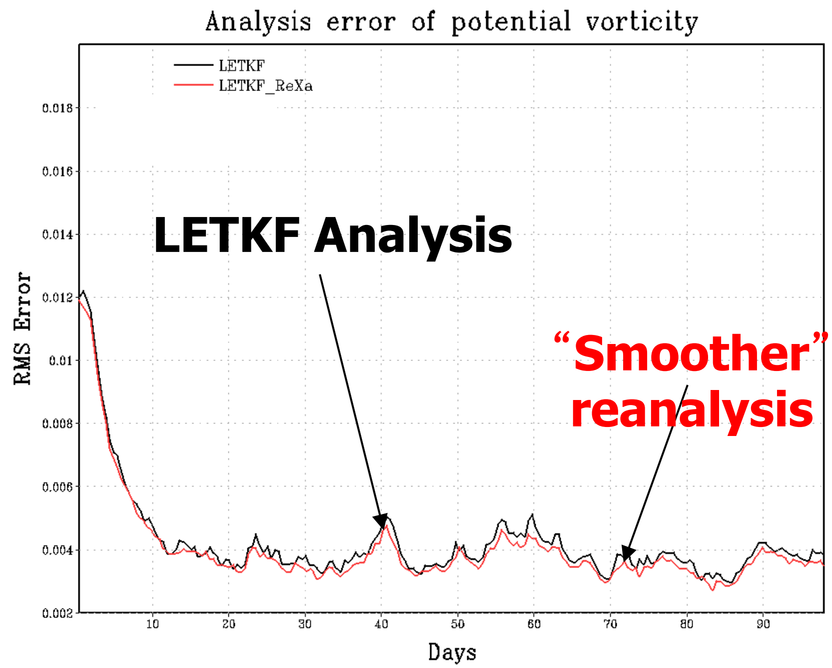
The no-cost smoother makes possible:

- ✓ Quasi Outer Loop (QOL)
- ✓ "Running in place" (RIP) for faster spin-up
- ✓ Use of future data in reanalysis
- ✓ Ability to use longer windows and nonlinear perturbations

No-cost LETKF smoother first tested on a QG model: it works…

LETKF analysis at time n
$$\overline{\mathbf{x}}_n^a = \overline{\mathbf{x}}_n^f + \mathbf{X}_n^f \overline{\mathbf{w}}_n^a$$

Smoother analysis at time n-1
$$\tilde{\mathbf{x}}_{n-1}^a = \overline{\mathbf{x}}_{n-1}^f + \mathbf{X}_{n-1}^f \overline{\mathbf{w}}_n^a$$



Very simple smoother: apply the final weights at the beginning of the window. It allows assimilation of future data, and assimilating data more than once.

Nonlinearities, "QOL" and "Running in Place"

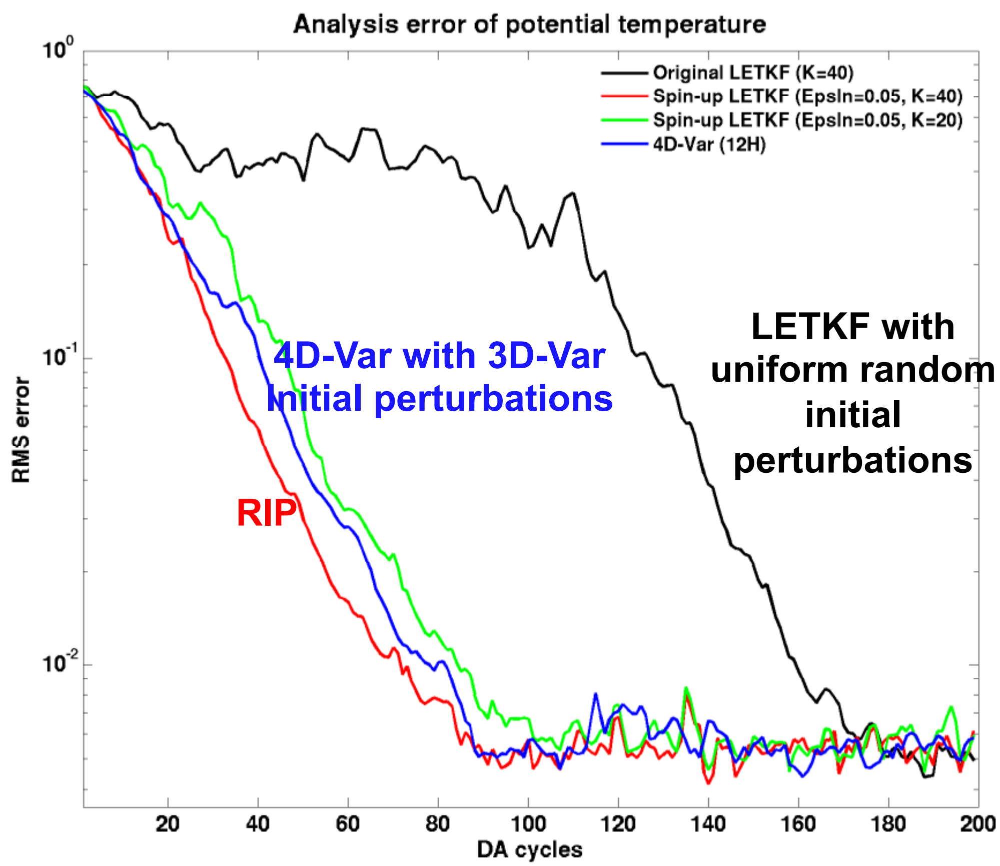
Quasi Outer Loop: use the final weights to correct only the <u>mean</u> initial analysis, keeping the initial perturbations. Repeat the analysis once or twice. It centers the ensemble on a more accurate nonlinear solution.

Lorenz -3 variable model RMS analysis error

	4D-Var	LETKF	LETKF +QOL	LETKF +RIP
Window=8 steps	0.31	0.30	0.27	0.27
Window=25 steps	0.53	0.68	0.47	0.35

"Running in Place" smooths both the analysis and the analysis error covariance and iterates a few times…
It can return to regular LETKF after spin-up.

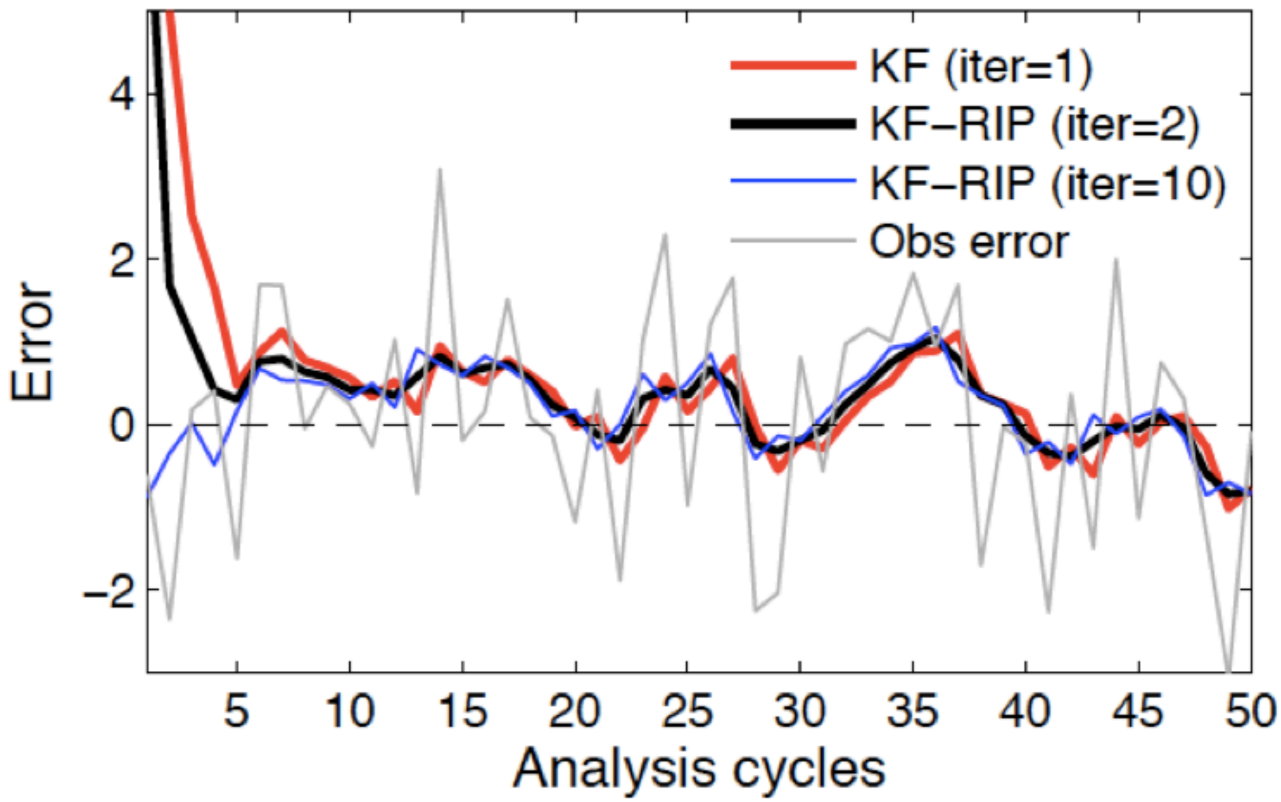
Running in Place: Spin-up with a QG model



RIP accelerates the EnKF spin-up (e.g., hurricanes, severe storms)

Spin-up depends on the initial perturbations, but RIP works well even with uniform random perturbations. RIP becomes even faster than 4D-Var (blue).

Why RIP works: Results with a Linear model

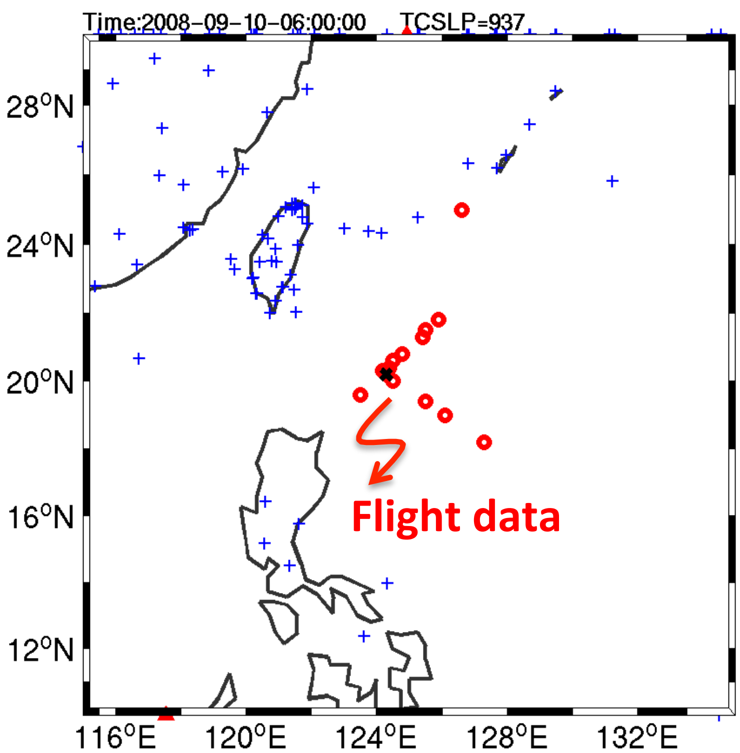


$$x_n = M(x_{n-1}) = x_{n-1} + \alpha$$

$$\sigma_n^2 = G(\sigma_{n-1}^2) = C\sigma_{n-1}^2$$

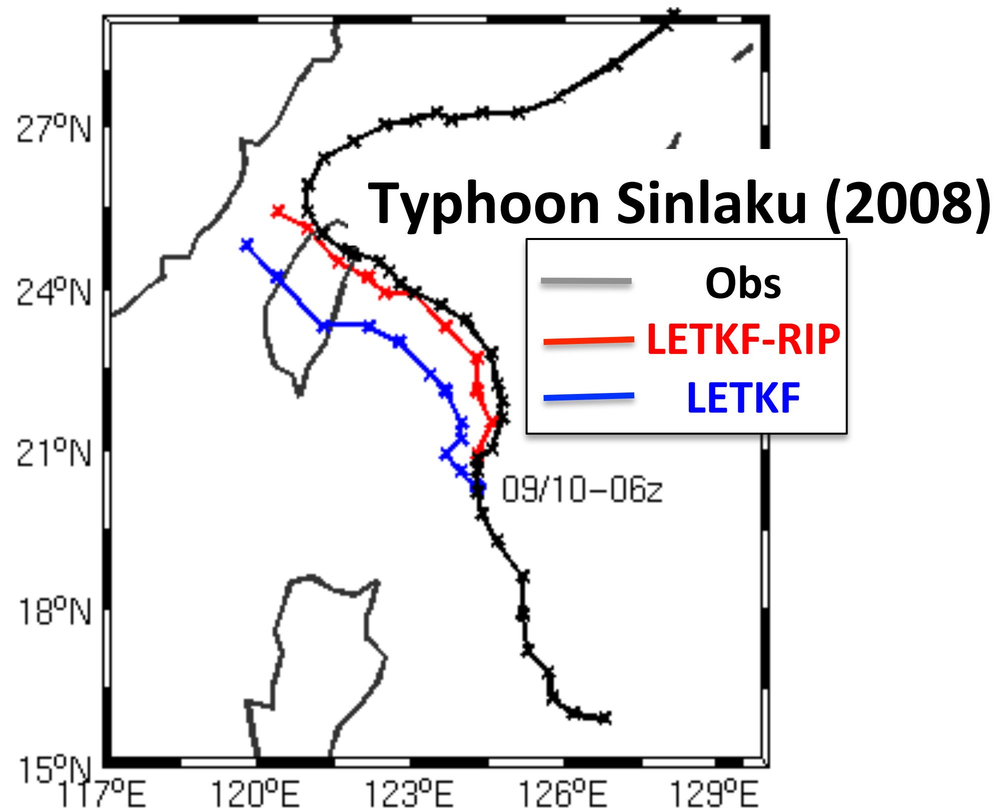
RIP adapts to using an observation N-times by dividing the spread by N: RIP converges to the regular optimal KF solution. The spin-up is faster and the analysis update is "softer" (in small steps) rather than in large steps.

LETKF-RIP with real observations (Typhoon Sinlaku, 2008)



SYNOP(+),SOUND(△), DROPSONDE(○), Typhoon center (X)

3-day forecast



RIP uses better the "limited observations"!

Courtesy of Prof. Shu-Chih Yang (NCU, Taiwan)

An application of LETKF-RIP to ocean data assimilation

Data Assimilation of the Global Ocean using 4D-LETKF, SODA (OI) and MOM2

Steve Penny's thesis defense

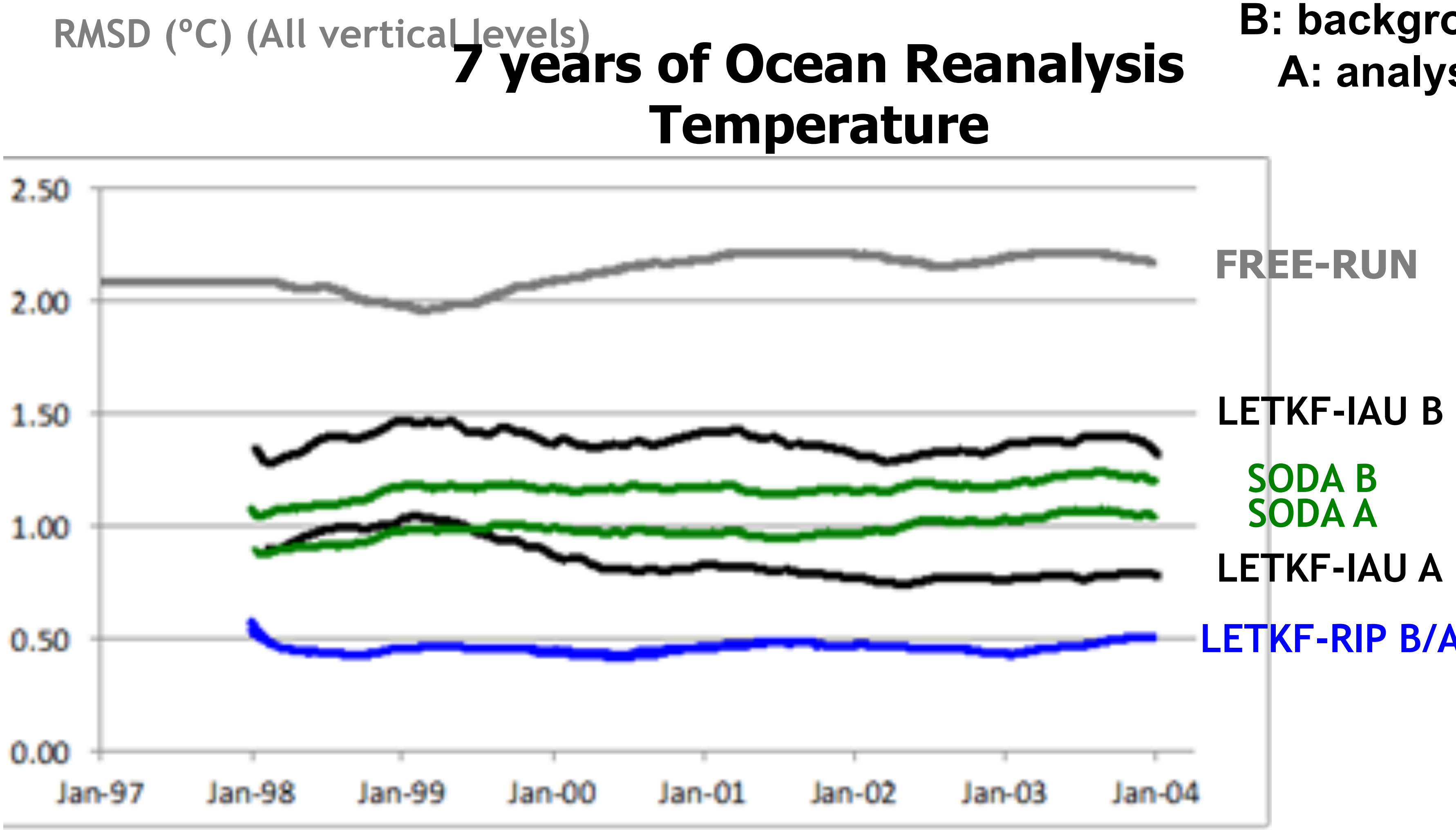
April 15, 2011

Advisors: E Kalnay, J Carton, K Ide, T Miyoshi, G Chepurin

Penny (now at UMD/NCEP) implemented the LETKF with either IAU or RIP and compared it with SODA (OI)

7 years of Ocean Reanalysis Temperature

B: background
A: analysis

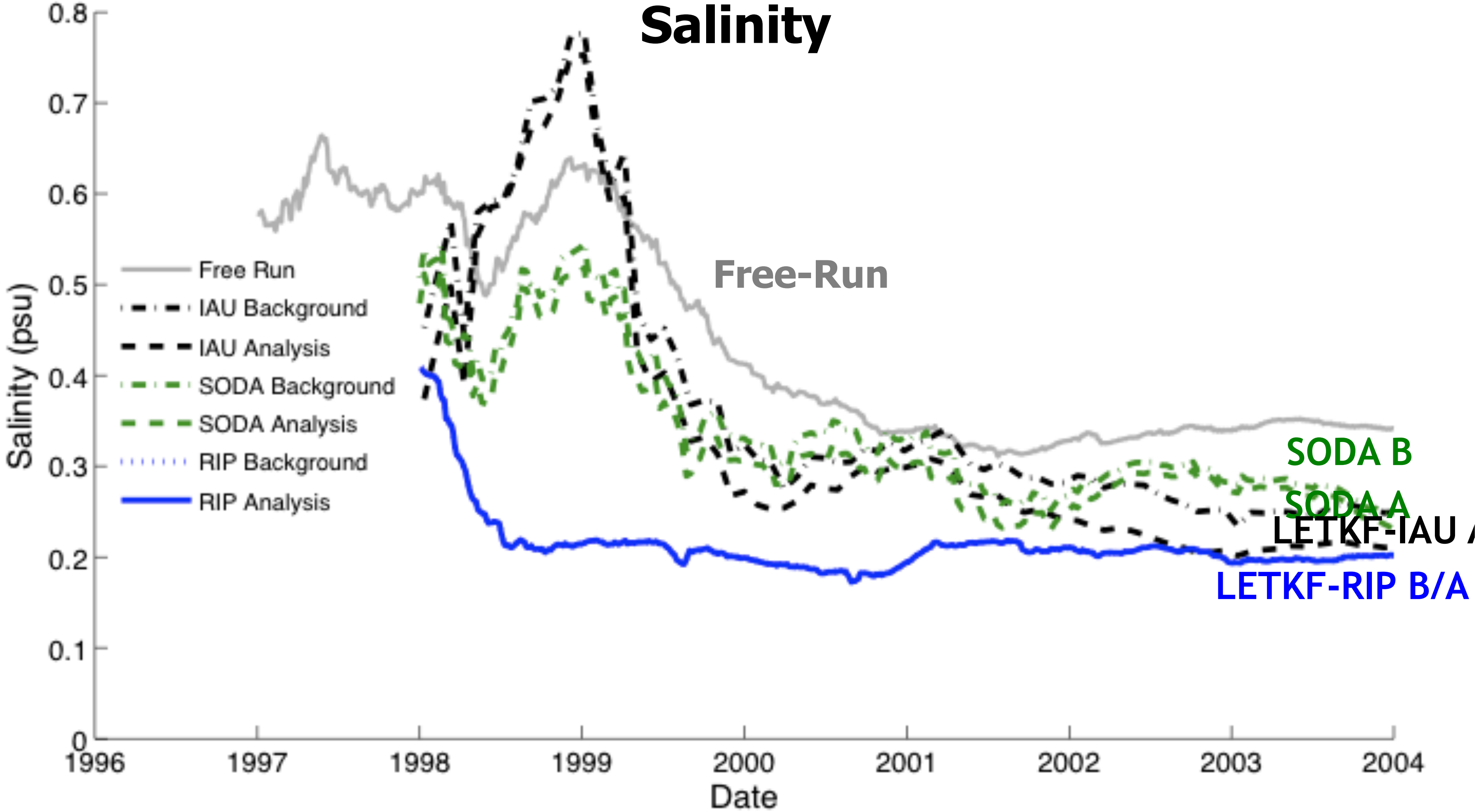


Global RMS(O-F) of Temperature (°C), 12-month moving average
LETKF (with IAU), SODA and LETKF with RIP

RMSD (psu) (All vertical levels)

B: background
A: analysis

7 years of Ocean Reanalysis Salinity



Global RMS(O-F) of Salinity (psu), 12-month moving average

LETKF (with IAU), SODA and LETKF with RIP

Promising new tools for the LETKF (2)

2. Effective assimilation of Precipitation (Guo-Yuan Lien, Eugenia Kalnay and Takemasa Miyoshi, 2013)

- Assimilation of precipitation has generally failed to improve forecasts beyond a day.
- A new approach deals with non-Gaussianity, and assimilation of both zero and non-zero precipitation.
- Rather than changing moisture to force the model to rain as observed, the LETKF changes the potential vorticity.
- The model now "remembers" the assimilation, so that medium range forecasts are improved.

How to transform precipitation *y* to a Gaussian y_{transf}

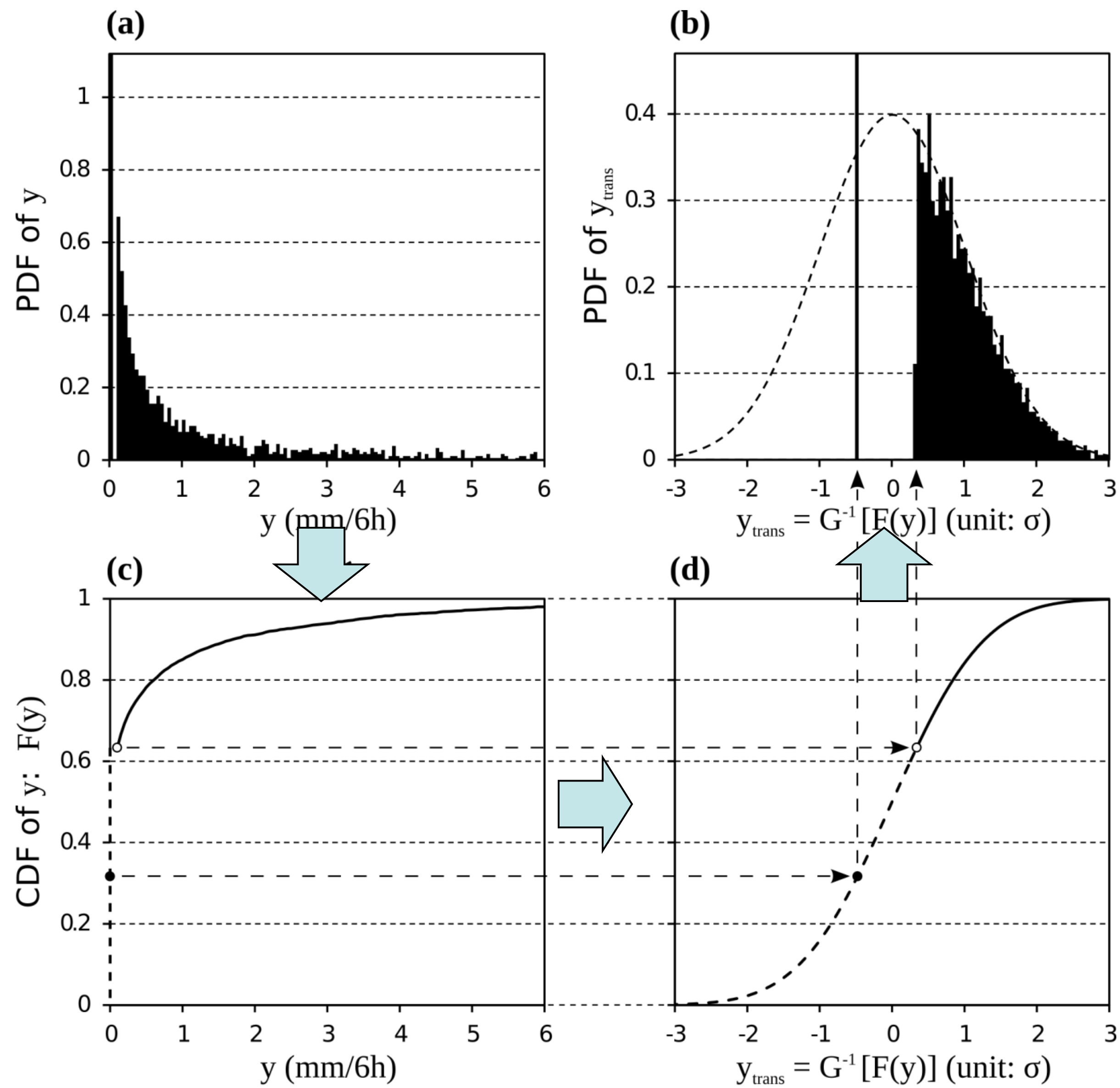
Start with pdf of *y*=rain at every grid point.

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

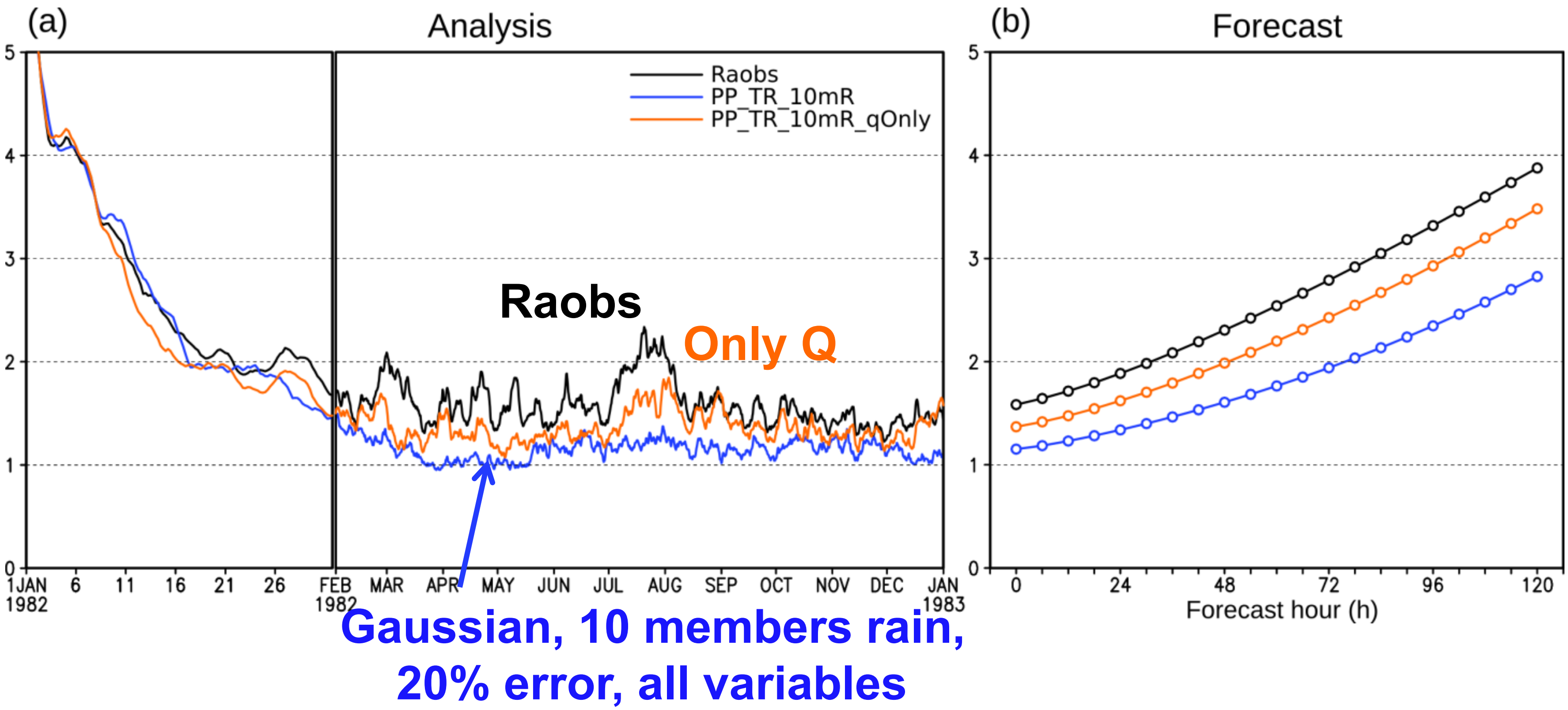
We assign all "no rain" to the median of the no rain CDF.

We found this works as well as more complicated procedures.

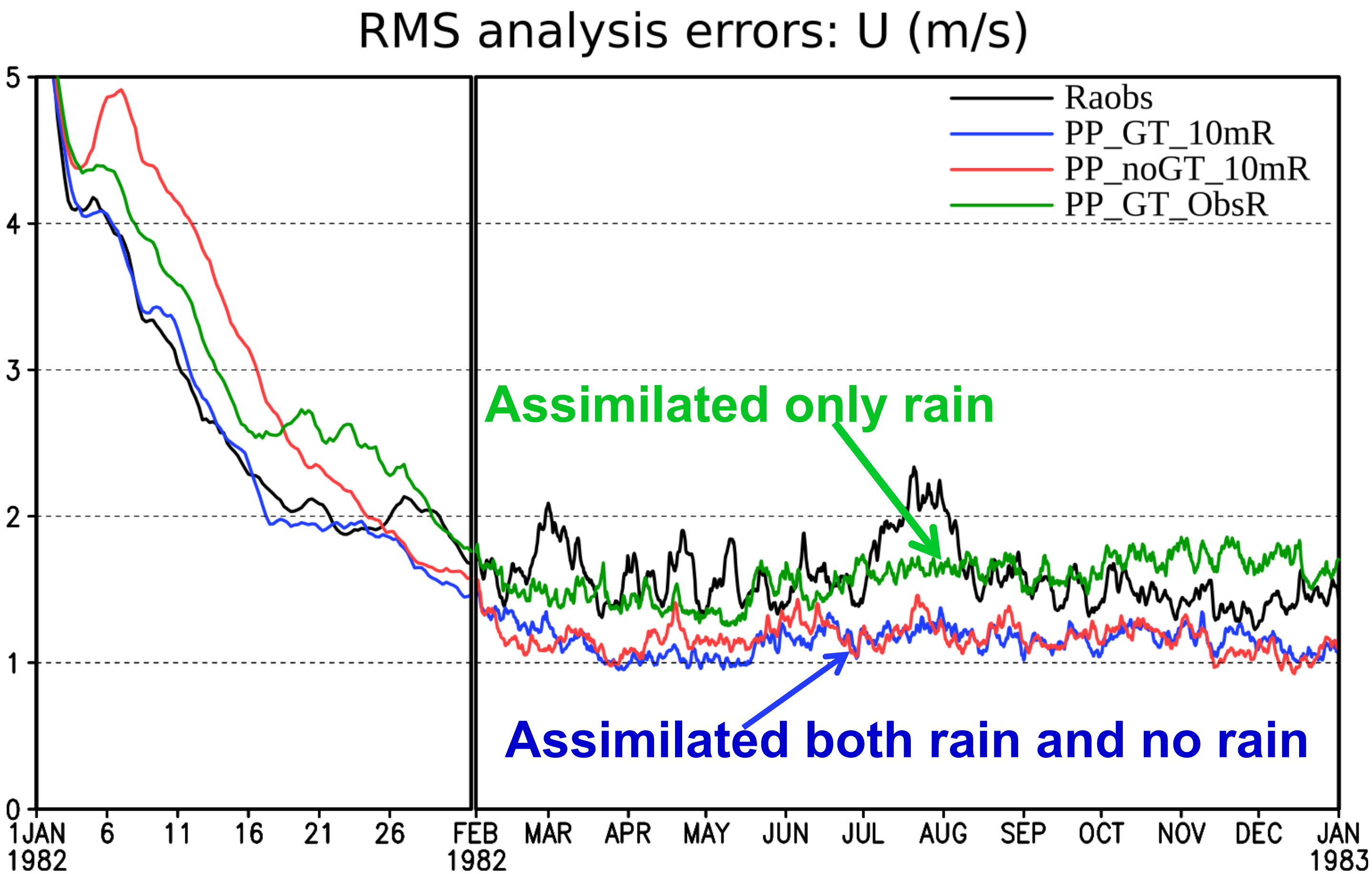
It allows to assimilate both rain and no rain.



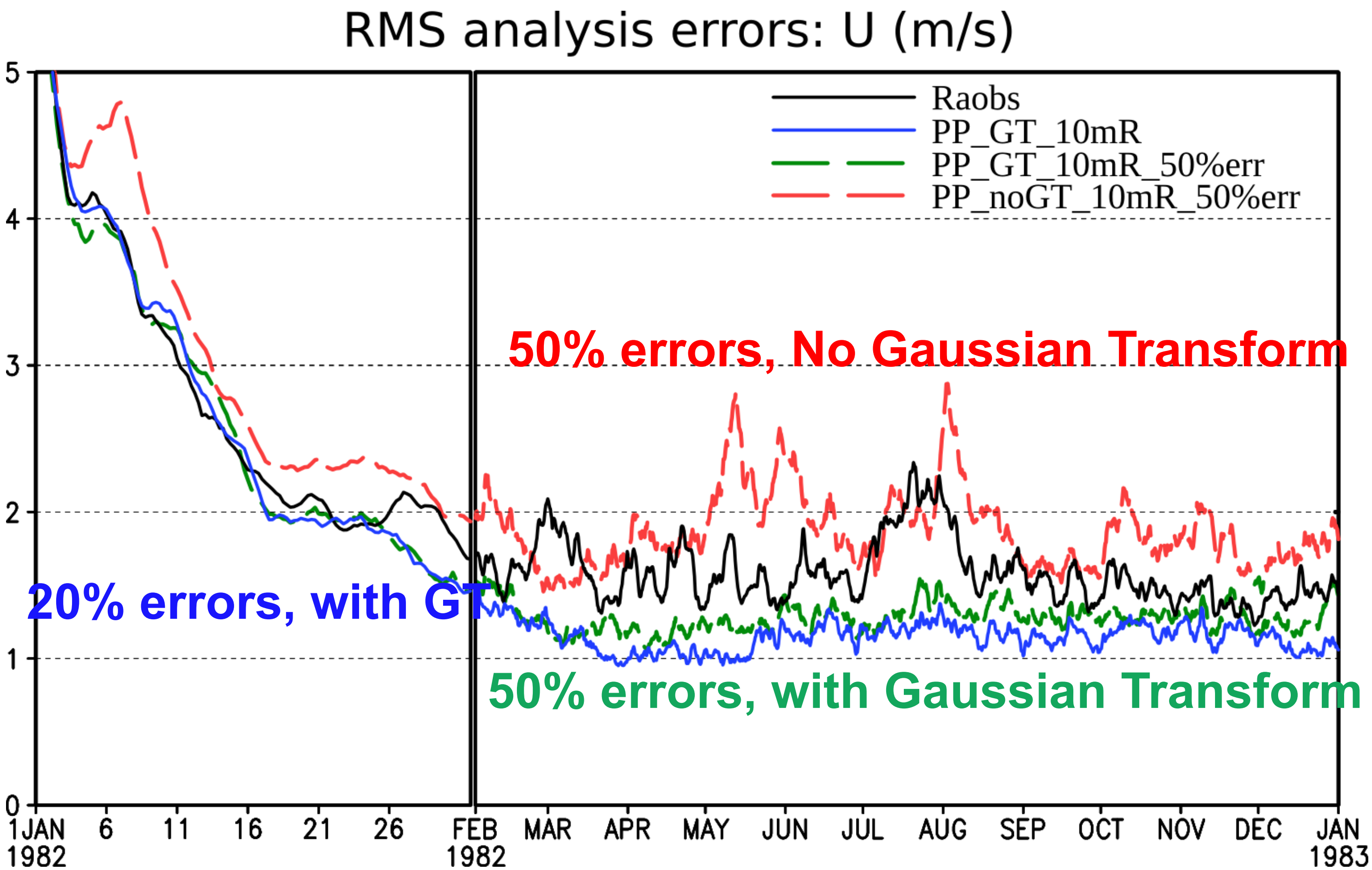
$$G^{-1}(x) = \sqrt{2}\, erf^{-1}(2x-1)$$



- 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!

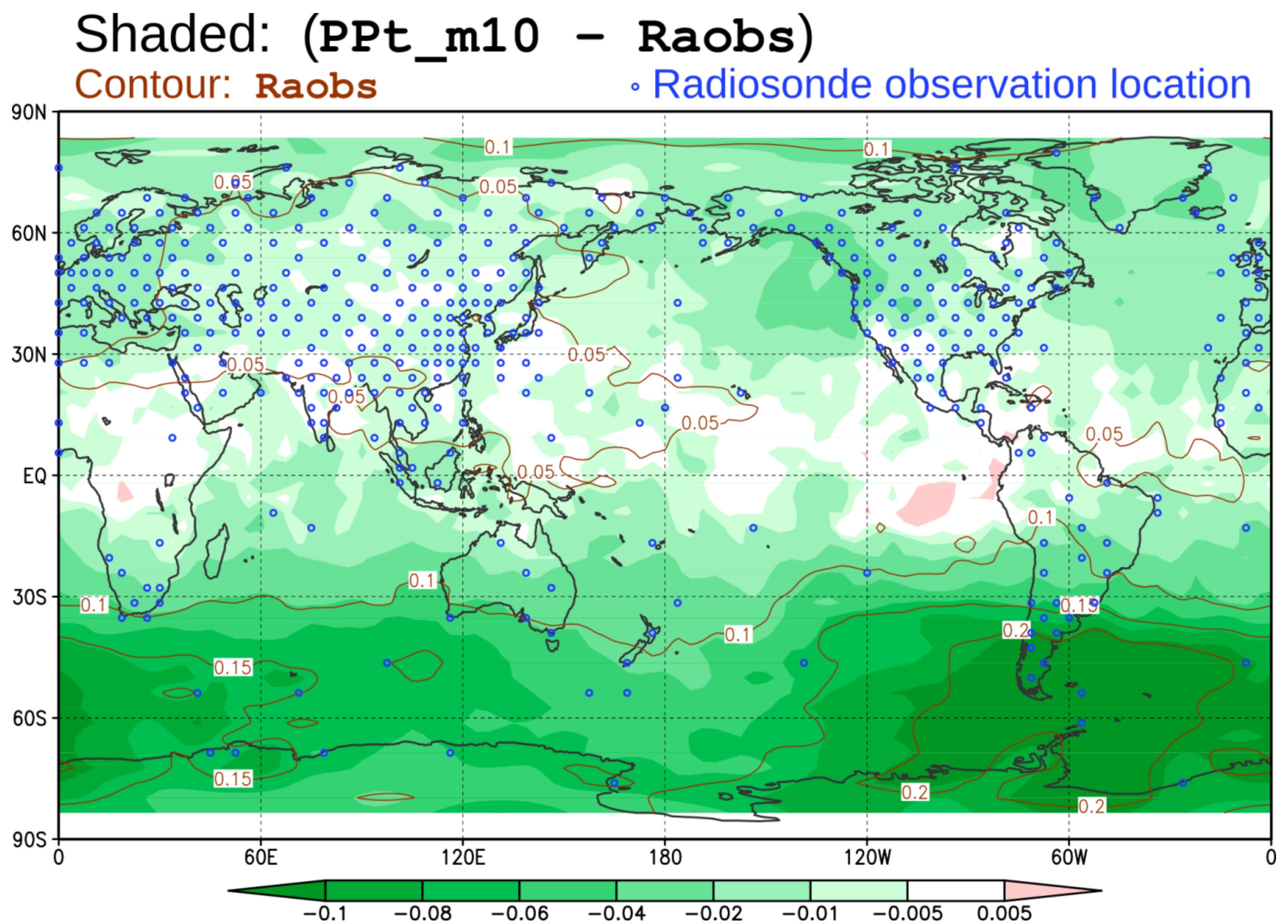


If we assimilate only rain the results are much worse!
We need to assimilate both rain and no rain!



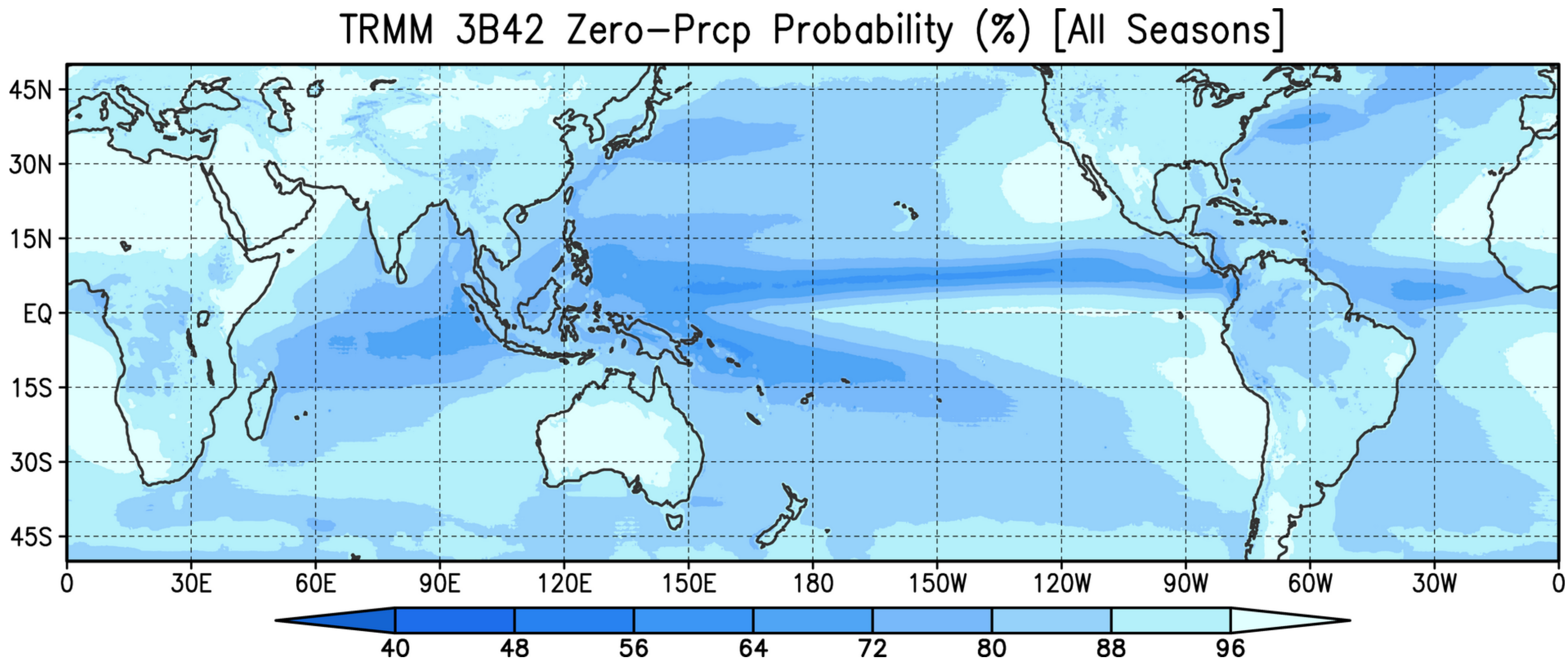
The impact of the Gaussian Transform is important with large observation errors (50% rather than 20%). The impact of GT50% is almost as good as GT20%.

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-LETKF



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

Promising new tools for the LETKF (3)

3. Forecast Sensitivity to Observations and Proactive QC

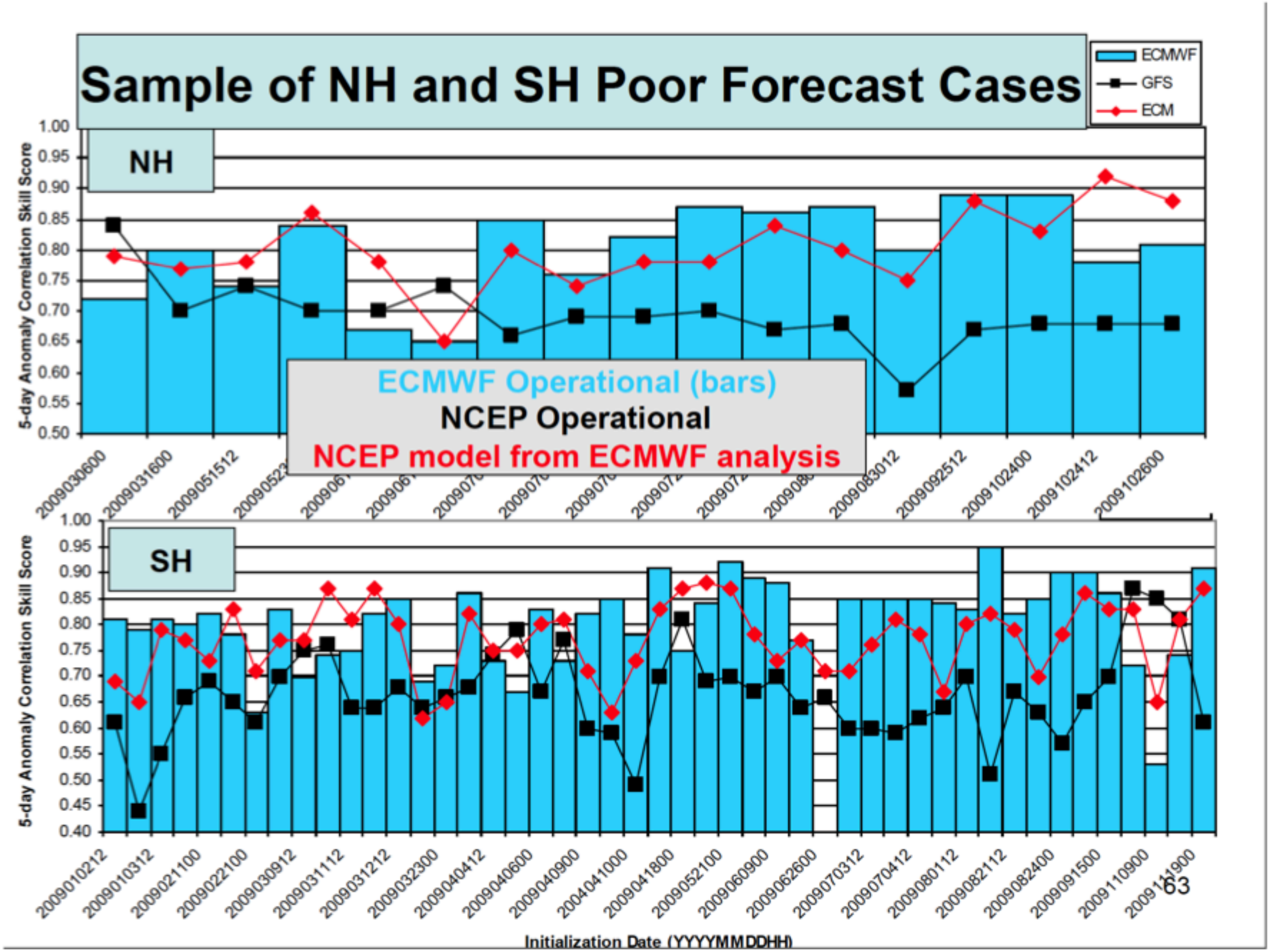
(with Y Ota, T Miyoshi, J Liu, and J Derber)

• A simpler, more accurate formulation 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 <u>all operational observations</u>.

• 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

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

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

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

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

Langland and Baker (2004), and Gelaro et al. 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):

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

With EnKF we can use the original equation without adjoint

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**

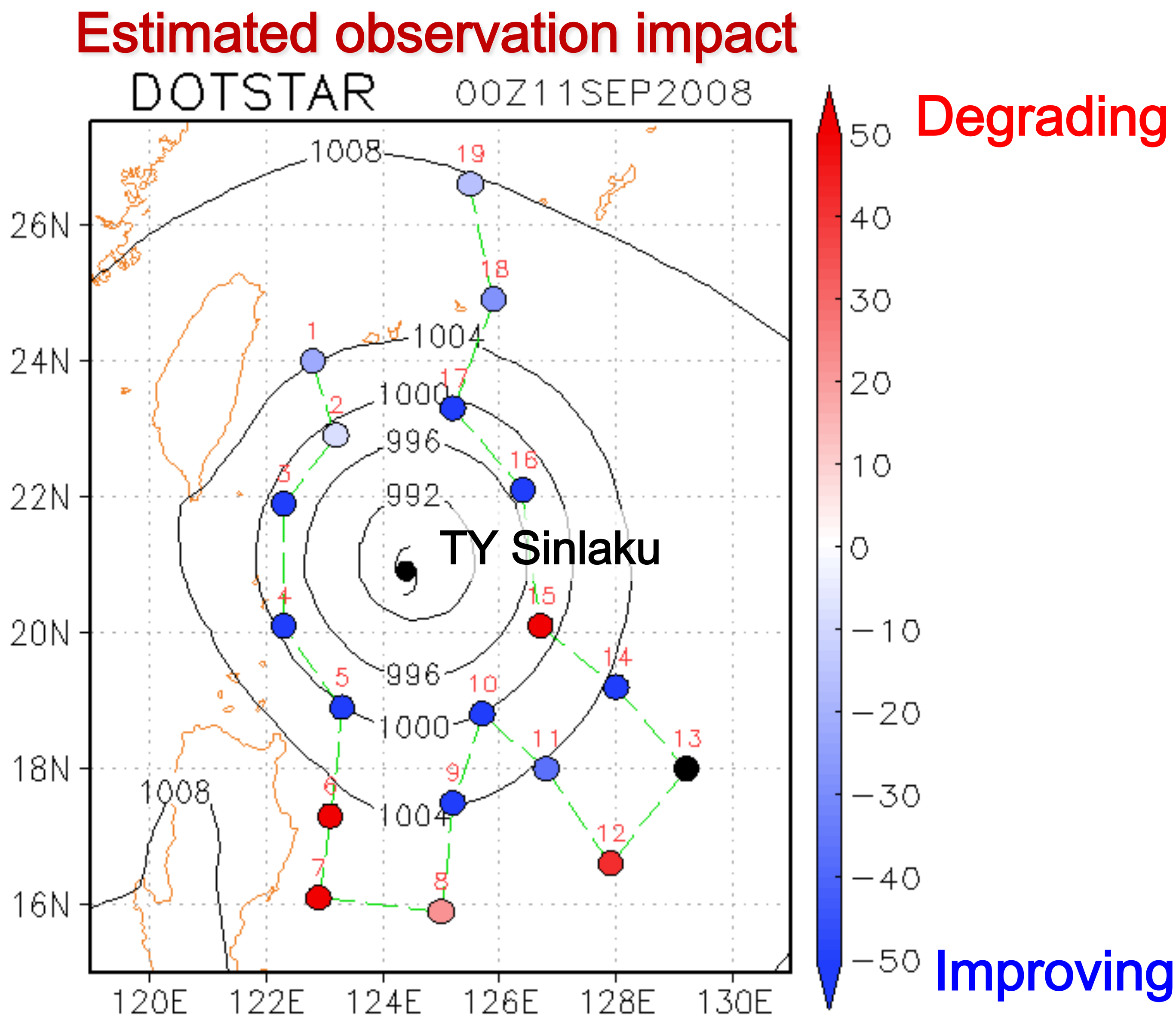
$$\mathbf{MK} = \mathbf{MX}^a (\mathbf{X}^{aT} \mathbf{H}^T) \mathbf{R}^{-1} / (K-1) = \mathbf{X}^f_{t|0} \mathbf{Y}^{aT} \mathbf{R}^{-1} / (K-1)$$

Thus,

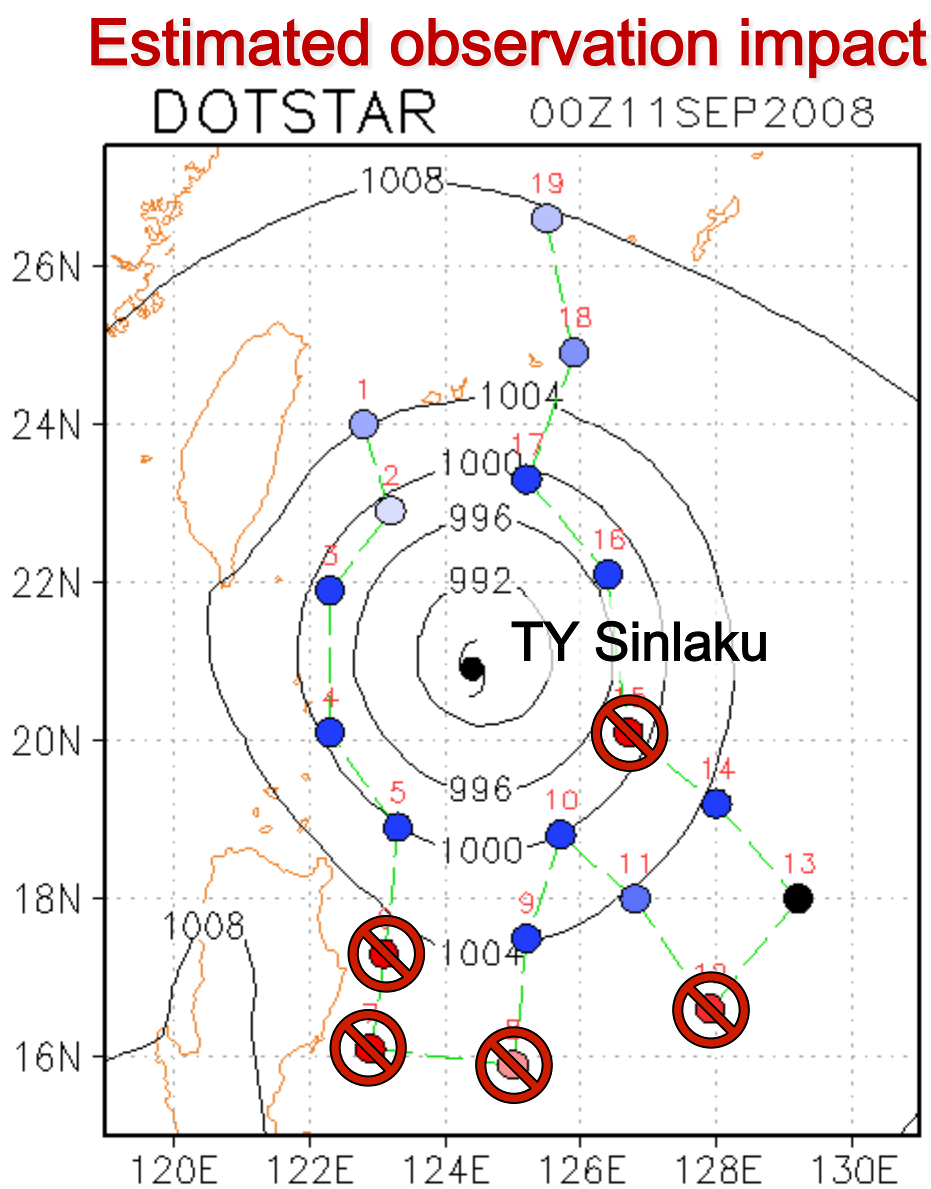
$$\Delta \mathbf{e}^2 = \left[\mathbf{MK}(\mathbf{y} - H(\mathbf{x}^b_{0|-6})\right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$
$$= \left[(\mathbf{y} - H(\mathbf{x}^b_{0|-6})\right]^T \mathbf{R}^{-1} \mathbf{Y}^a_0 \mathbf{X}^{fT}_{t|0} (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) / (K-1)$$

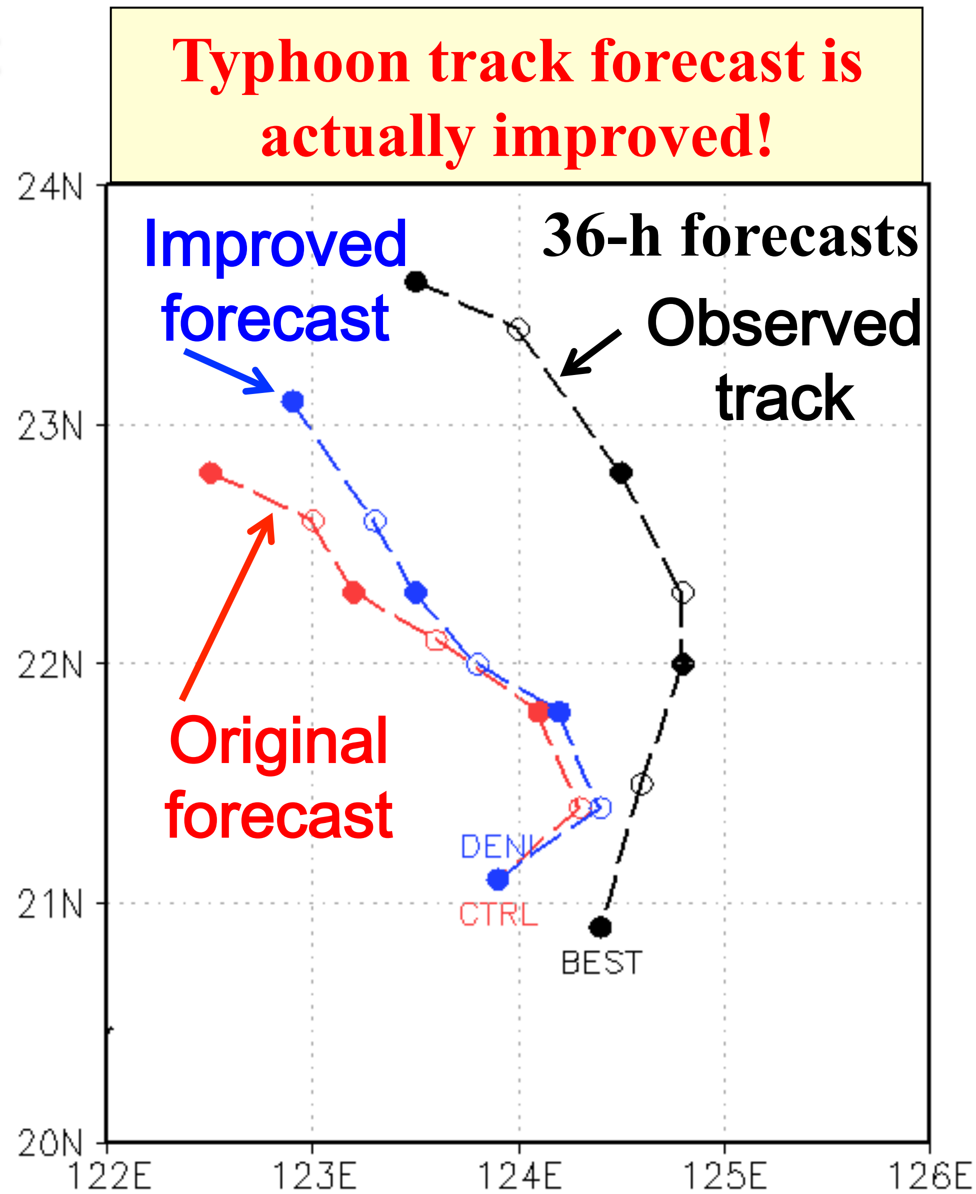
This uses the available nonlinear forecast ensemble products.

Impact of dropsondes on a Typhoon
(Kunii et al. 2012)



Denying negative impact data improves forecast!





Ota et al. 2012: Applied EFSO to NCEP GFS/ EnSRF using all operational observations. Determined regional 24hr "forecast failures"

- Divide the globe into 30x30° 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 skill "dropout"

24-hr forecast error correction (Ota et al., 2013)

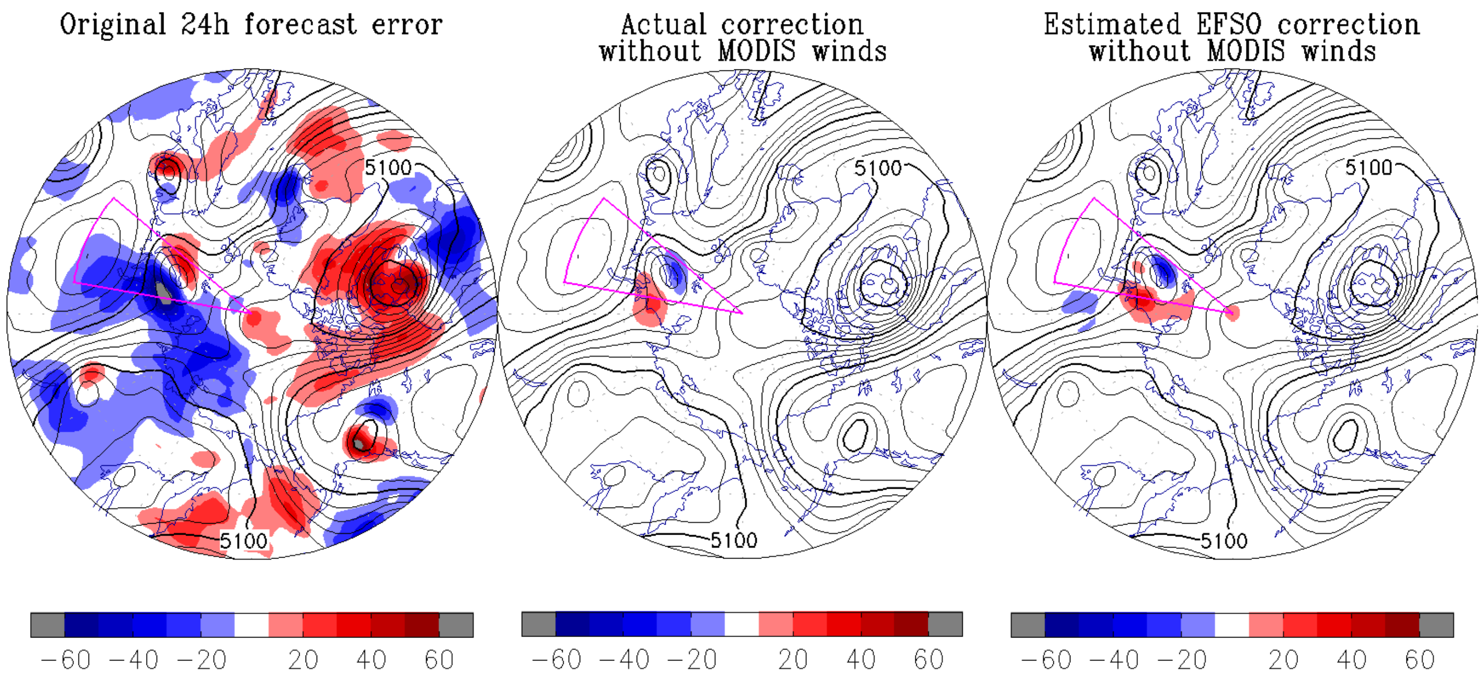
- identified 7 cases of large 30°x30° regional errors,
- rerun the forecasts denying bad obs.
- the forecast errors were substantially reduced
- this could be applied to do "proactive QC"

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%

MODIS

“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.

Promising new tools for the LETKF (4)

4. Estimation of surface fluxes as evolving parameters

(Kang et al., 2011, Kang et al., 2012)

- Important for the carbon cycle
- Surface fluxes of heat, moisture, and momentum
- Eventually for coupled data assimilation

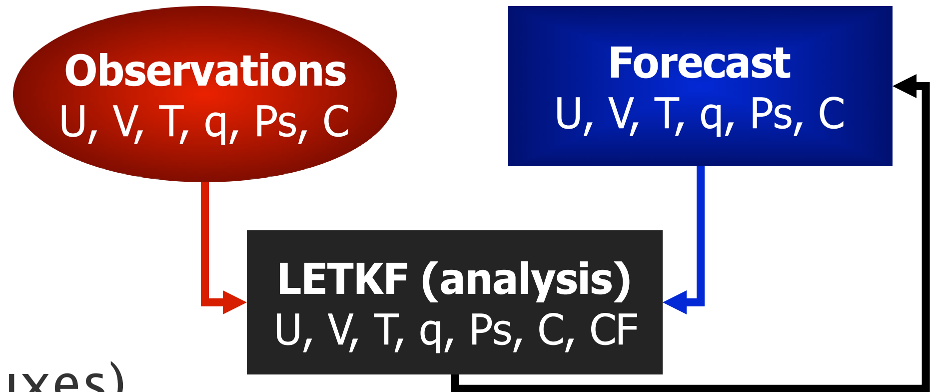
Parameter estimation in EnKF

- State vector augmentation

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

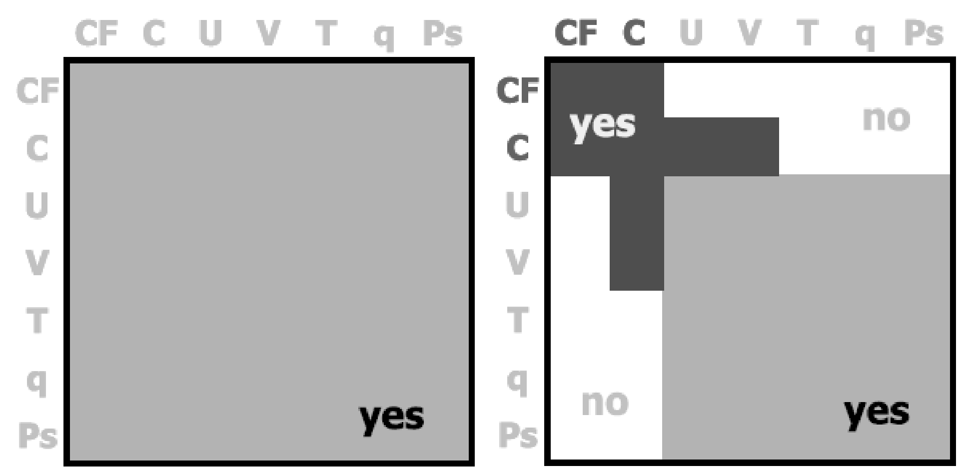
: model state vector (U, V, T, q, Ps, C)

: surface CO_2 flux



  - Append **CF** (surface CO_2 fluxes)
  - Update **CF** as part of the data assimilation process
  - **Multivariate** analysis with a localization of the variables (Kang et al., 2011)

Schematic plots of background error covariance matrix P^b ➔ without "variable localization" (left) and with it (right)



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

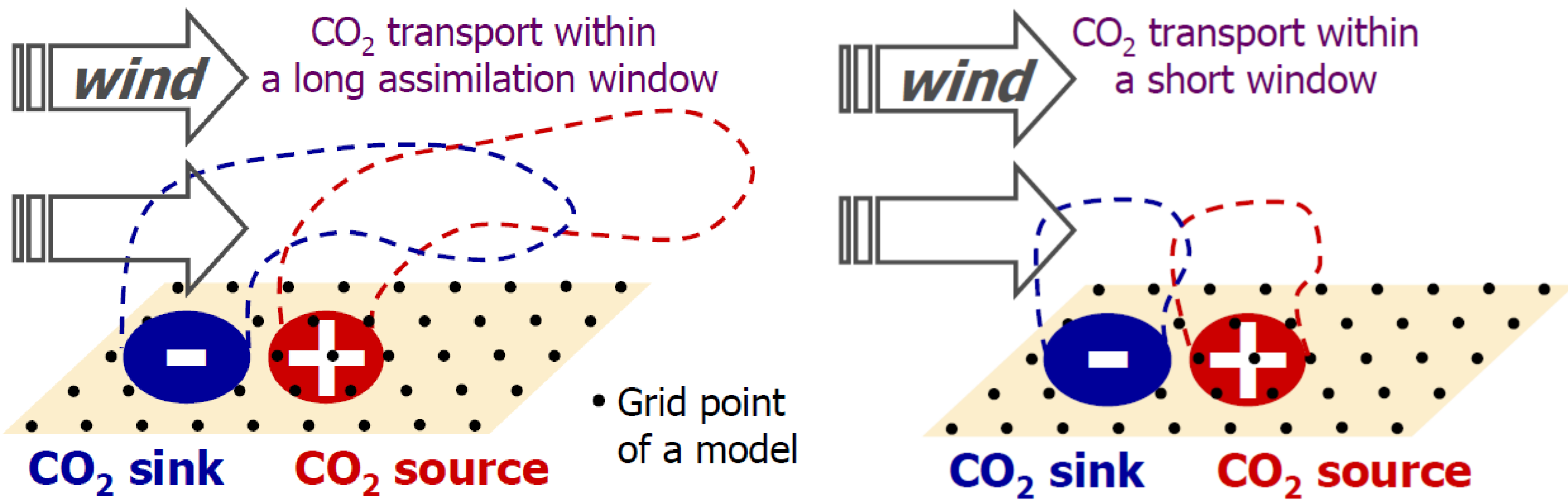
LETKF-C

- Carbon cycle data assimilation within LETKF (Kang et al., JGR, 2011, 2012)
  - Simultaneous analysis of meteorological and carbon variables
  - "Localization of Variables" reduces sampling errors
  - Advanced inflation methods
  - Vertical localization of column mixing CO_2 observations
  - Short (6-hour) assimilation window
    - Many of CO_2 inversion groups adopt much **longer window lengths** (weeks ~ months)
      - Started in the 1980's when there were only tens or hundreds of ground-based observations on the globe
      - CO_2 is an inert gas that stays long in the atmosphere so that the atmospheric CO_2 has quite long memory of CF.

    ➔ ***We have satellite observations of CO_2*** *(e.g. AIRS, GOSAT, OCO-2)*

    ➔ ***The long memory can be useful only if we can keep track of CO2 flow***

Assimilation window in LETKF-C

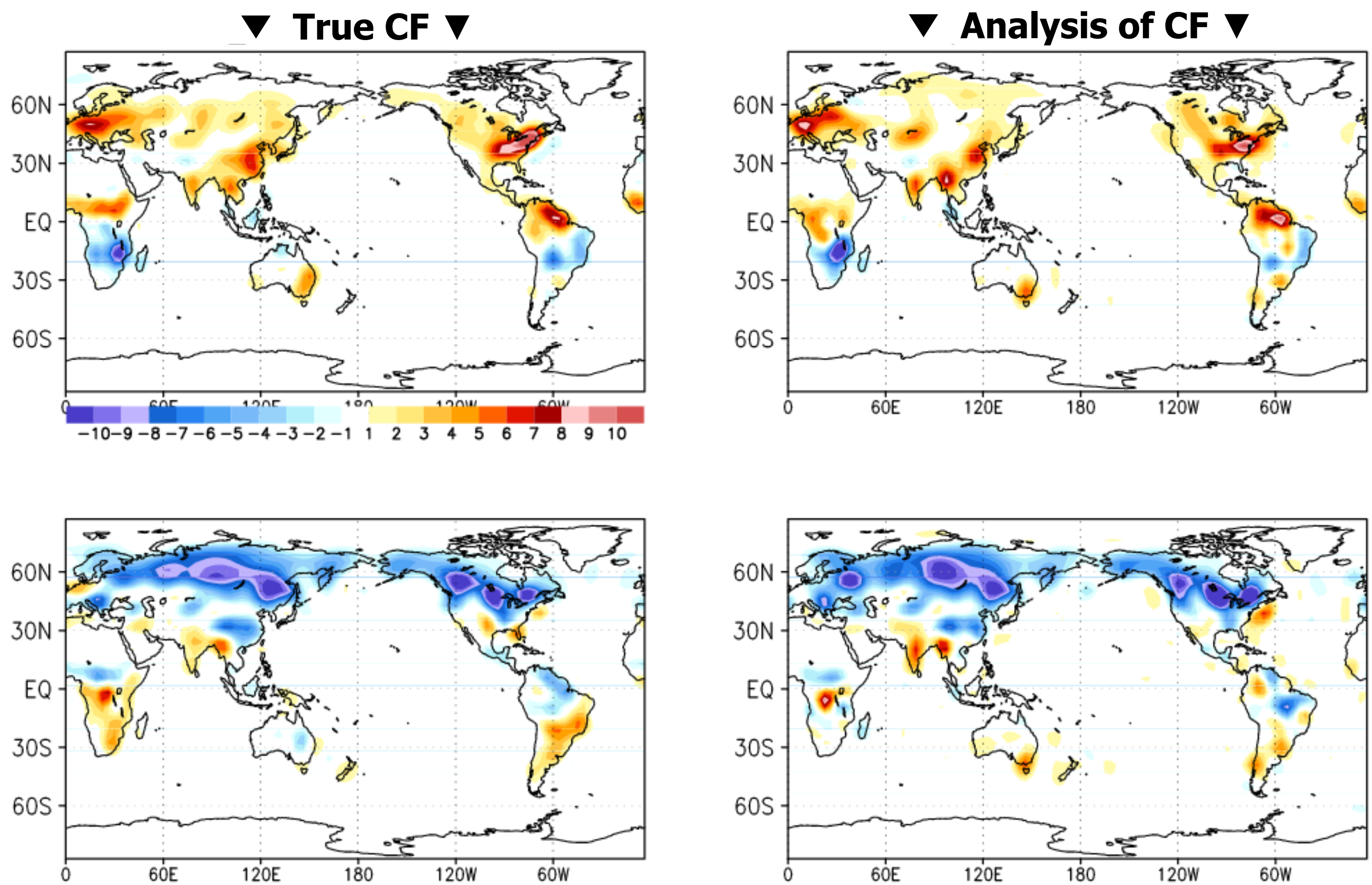


- CO_2 data assimilation system
  - A short assimilation window reduces the attenuation of observed CO_2 information because the analysis system can use the strong correlation between C and CF before the transport of atmospheric CO_2 **blurs out** the essential information of surface CO_2 forcing
  - 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

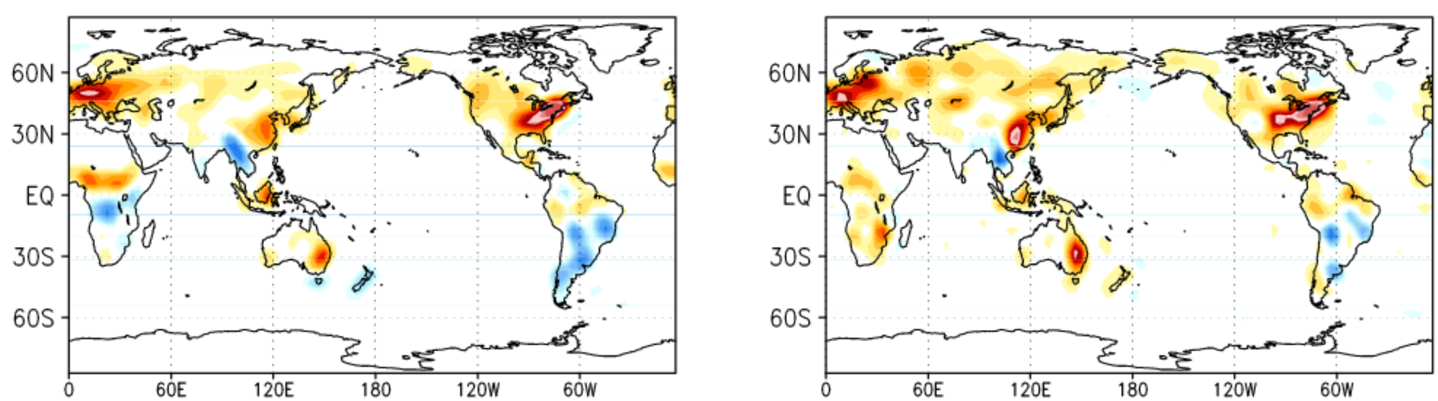
00Z01APR ▶
After three months of DA

00Z01AUG ▶
After seven months of DA



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

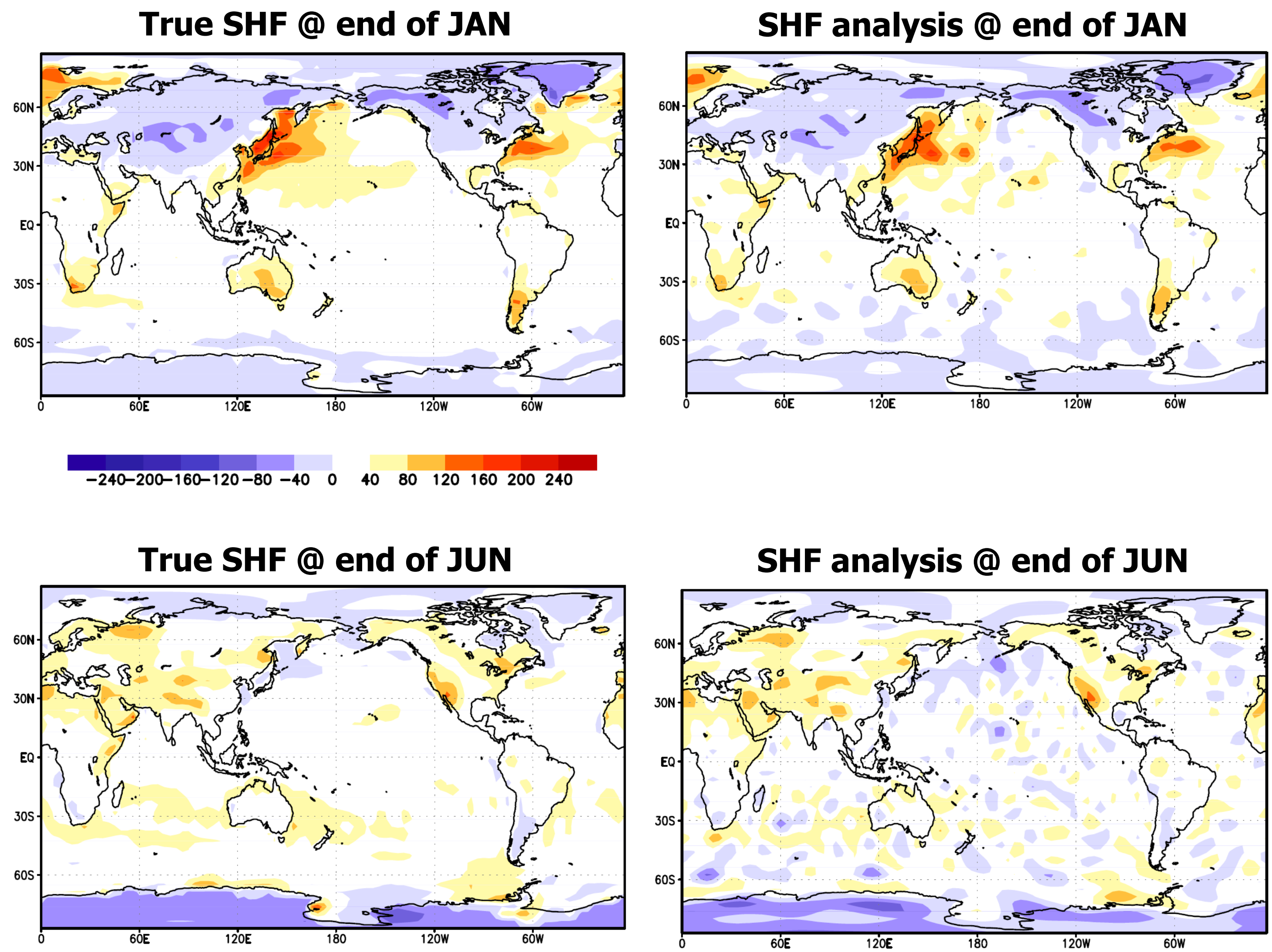
00Z01JAN ▶
After one year of DA



Surface Heat and Moisture Fluxes

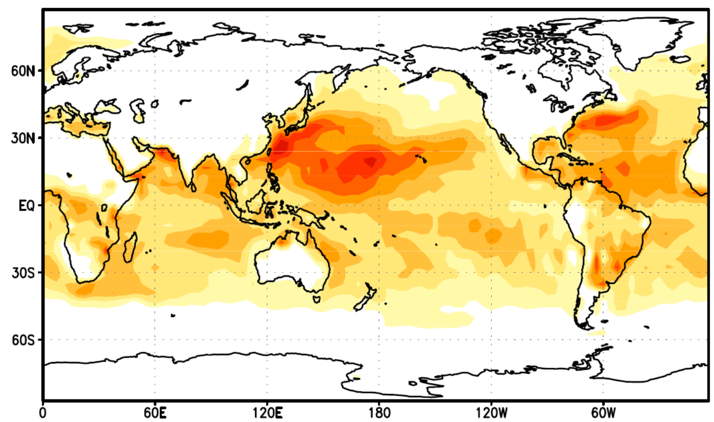
- Can we estimate **surface moisture/heat fluxes** by assimilating atmospheric moisture/temperature observations? *We can use the same methodology!*
- OSSEs
  - Nature: SPEEDY (perfect model)
  - Forecast model: SPEEDY with **persistence forecast of Sensible/Latent heat fluxes (SHF/LHF)**
  - Observations: conventional observations of (U, V, T, q, Ps) and AIRS retrievals of (T, q)
  - Analysis: U, V, T, q, Ps + SHF & LHF
- Fully multivariate data assimilation
- Adaptive multiplicative inflation + additive inflation
- Initial conditions: random (*no a-priori information*)

Results: SHF (perfect wind stress model)

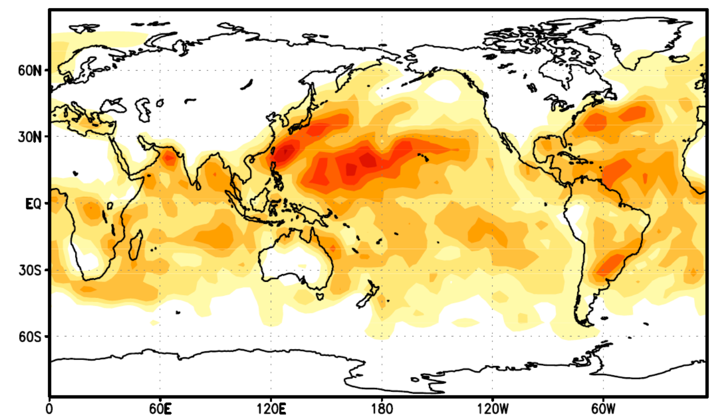


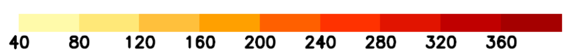
Results: LHF (perfect wind stress model)

True LHF @ end of JAN

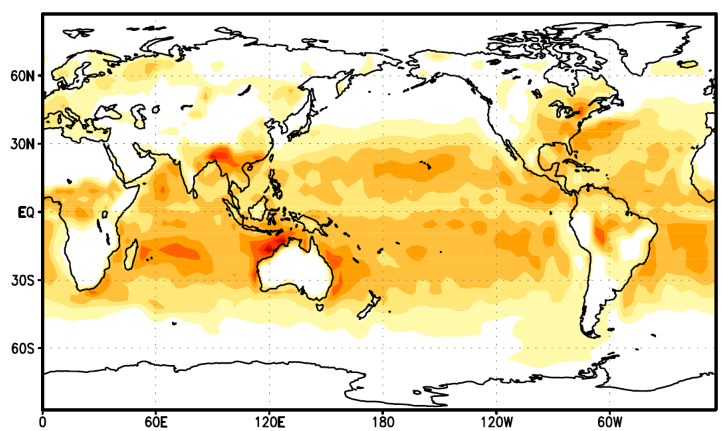


LHF analysis @ end of JAN

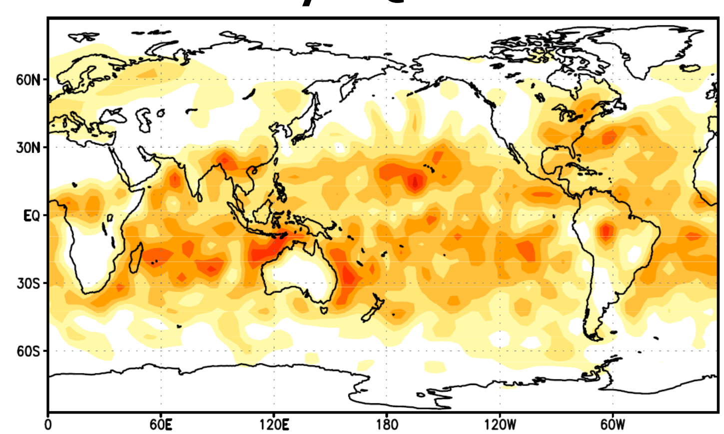




True LHF @ end of JUN



LHF analysis @ end of JUN



Summary

- We have shown the feasibility of simultaneous analysis of meteorological and carbon variables within LETKF framework through simulation experiments.
- The system LETKF-C has been tested in a intermediate-complexity model SPEEDY-C with excellent results.
  - Multivariate data assimilation with "localization of the variables" (Kang et al. 2011)
  - Advanced data assimilation methods for CO_2 flux estimation have been explored (Kang et al. 2012)
  - A short window is better than a long window.
- We are implementing the LETKF-C to NCAR CAM 3.5 model and real observations