# Recent advances in EnKF: Running in Place, Assimilation of Rain, Ens Fcast Sens to Obs, Coupled Data Assimilation

Shu-Chih Yang, Ji-Sun Kang, Takemasa Miyoshi, Steve Penny, Guo-Yuan Lien, Daisuke Ota…

Eugenia Kalnay

University of Maryland

UMD Weather-Chaos Group: **Kayo Ide**, **Brian Hunt**, and students (Travis Sluka, Yan Zhou, Adrienne Norwood, Erin Lynch, Yongjing Zhao)

ECMWF, 25/10/13

# 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)
- 7-years of Ocean Reanalysis (Penny, 2011, Penny et al., 2013)
- Very good results!

# 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.
- For perfect model experiments, the model now “remembers” the assimilation, so that that medium range forecasts are improved.
- Starting assimilation of real precipitation.

# Promising new tools for the LETKF(3)

**3. Forecast Sensitivity to Observations and "proactive QC"**

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

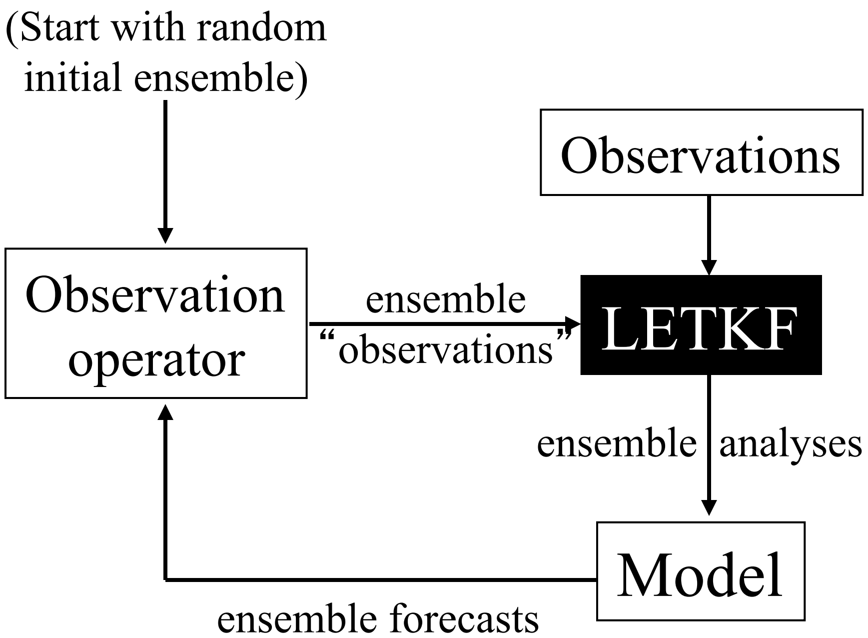
- A simpler, more accurate formulation for the Ensemble Forecast Sensitivity to Observations (EFSO, Kalnay et al., 2012, Tellus).
- Ota et al., 2012 tested it with the NCEP EnSRF-GFS operational system using <u>all operational observations</u>.
- Should allow identifying "bad observations" after 6hr, and then repeat the data assimilation without them: "proactive QC".

**4. Ensemble Singular Vectors** (Yang and Kalnay)

- Promising for additive inflation

**5. Coupled ocean-atm data assim., new hybrid** (Penny et al)

# Local Ensemble Transform Kalman Filter (Ott et al, 2004, Hunt et al, 2004, 2007) (a square root filter)

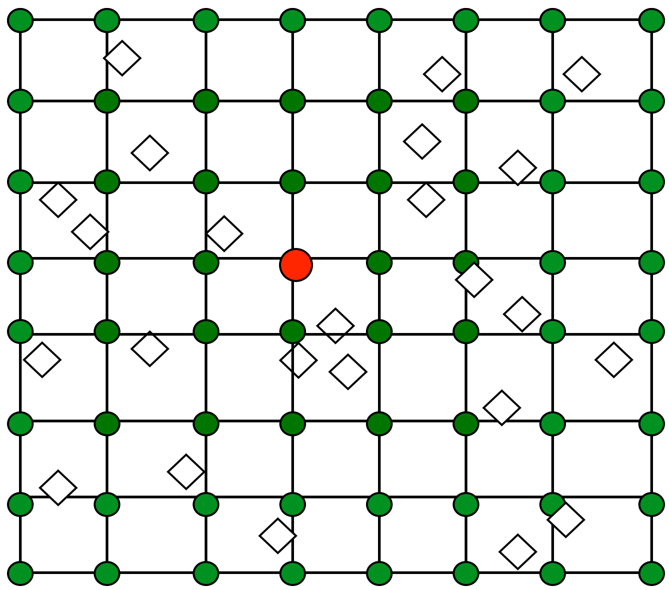



- Model independent (black box)
- Obs. assimilated simultaneously at each grid point
- 100% parallel
- No **adjoint** needed
- 4D LETKF extension
- Computes the **weights** for the ensemble forecasts explicitly

# Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

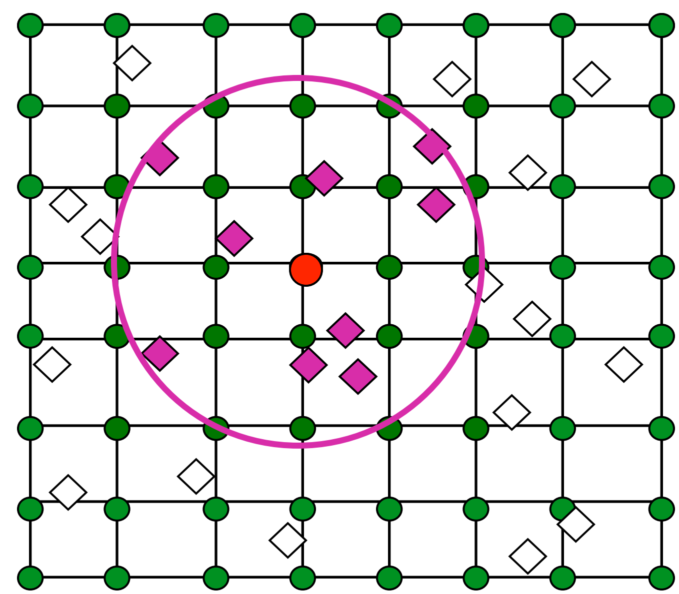


# Localization based on observations

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:

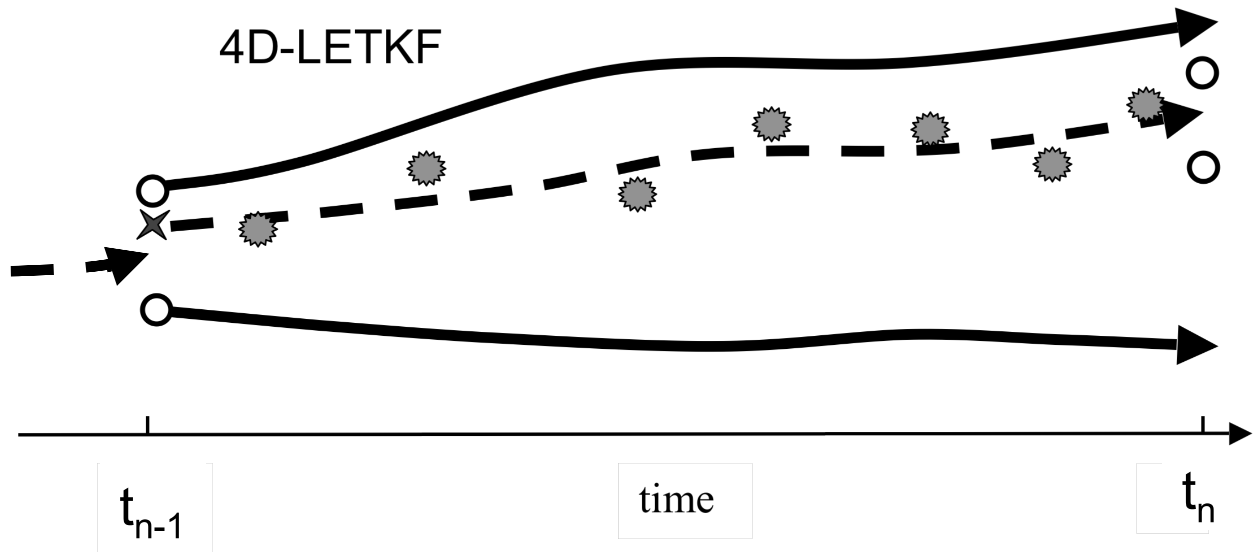
$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1};\ \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

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.**

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



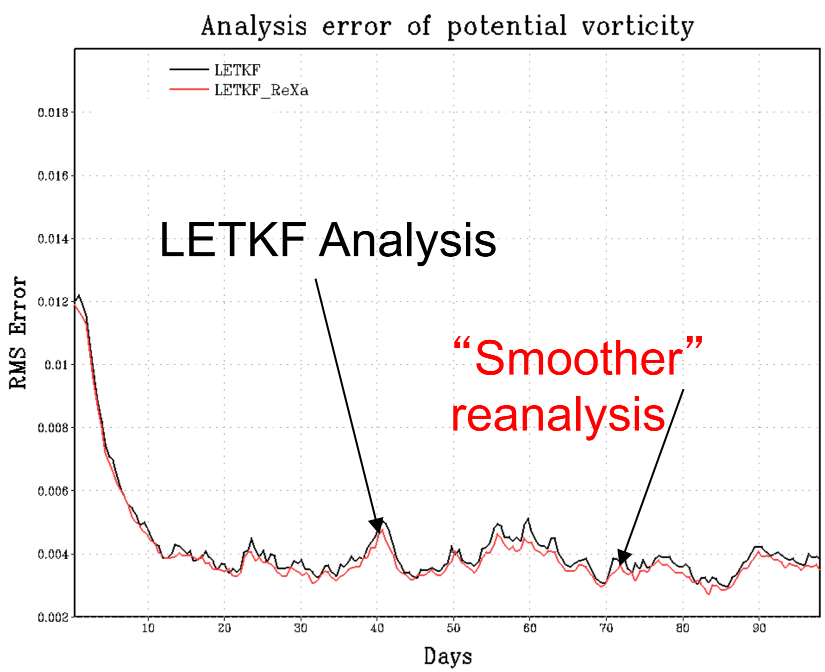

The no-cost smoother makes possible:

- ✓ 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 $n$ $\quad \overline{\mathbf{x}}_n^a = \overline{\mathbf{x}}_n^f + \mathbf{X}_n^f \overline{\mathbf{w}}_n^a$

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




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”

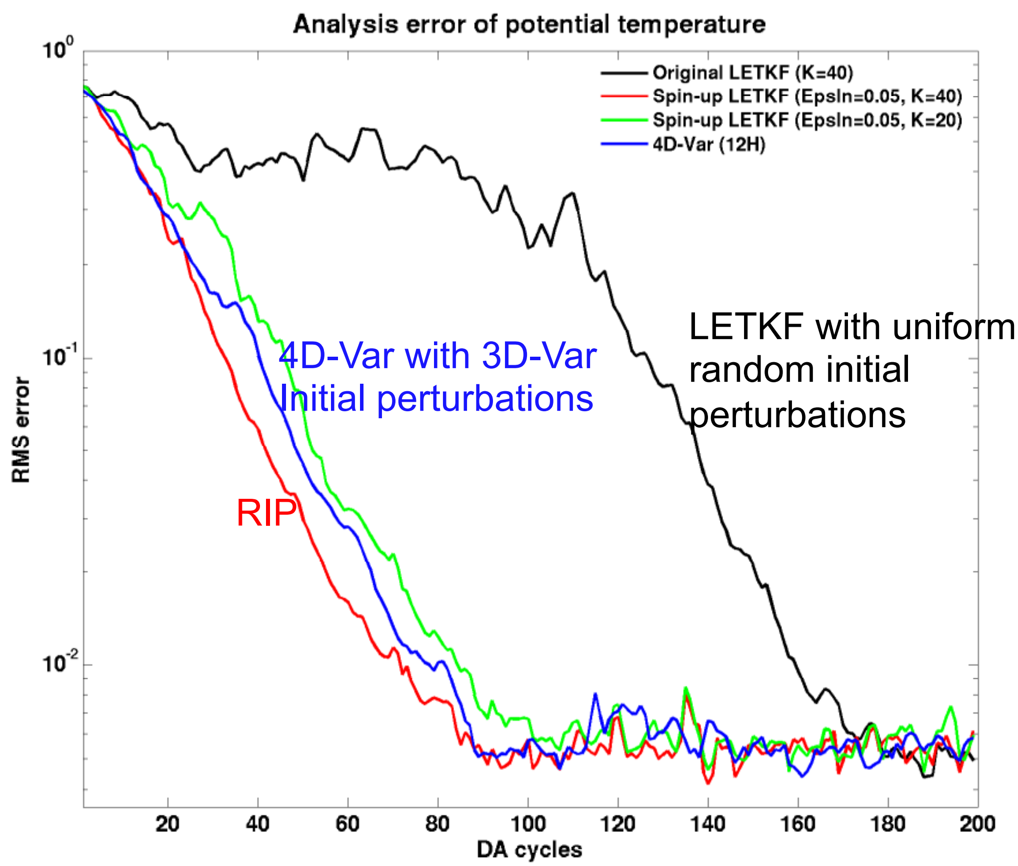
**Quasi Outer Loop: It centers the ensemble on a more accurate nonlinear solution.**

“**Running in Place**” smoothes both the **analysis** and the **analysis error covariance** and iterates a few times…

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** |

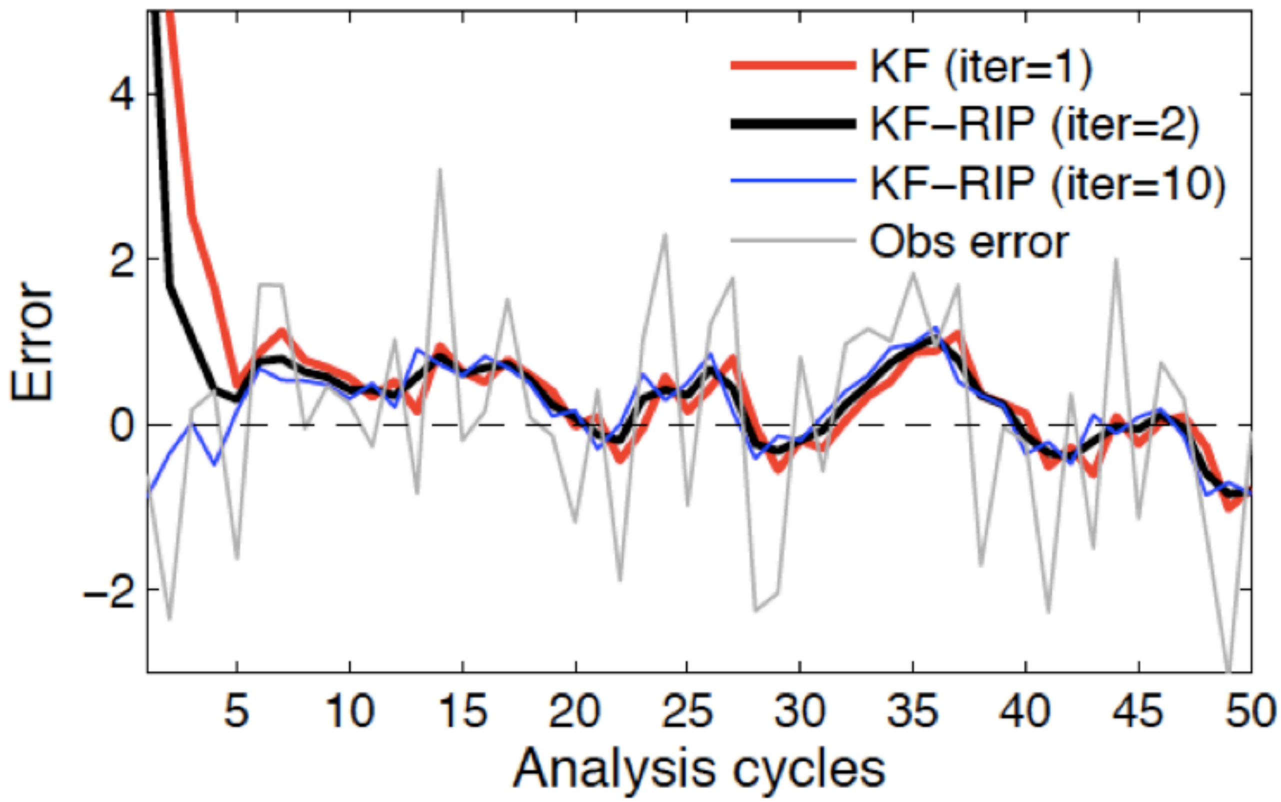
# Running in Place: Spin-up with a QG model




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

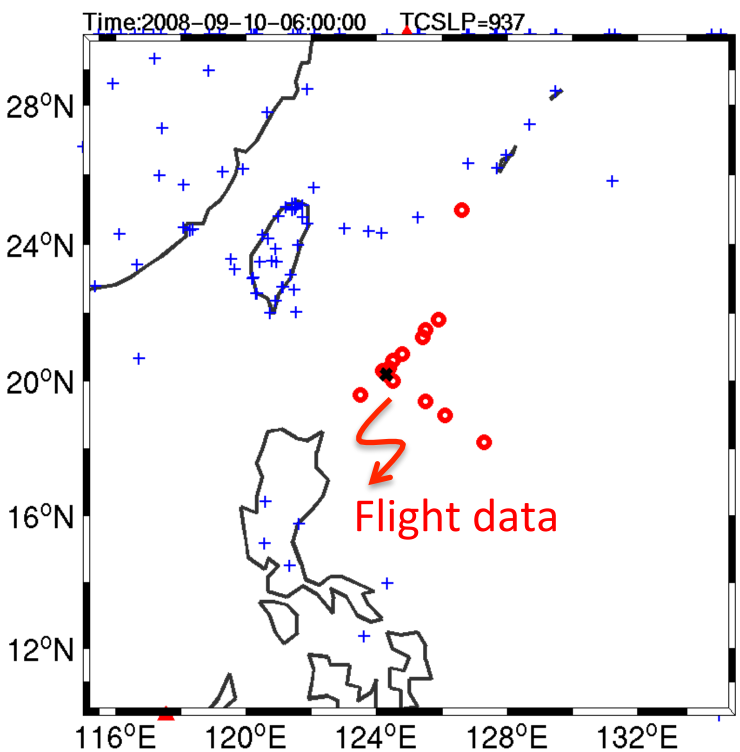



$$x_n = M(x_{n-1}) = x_{n-1} + \alpha$$

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

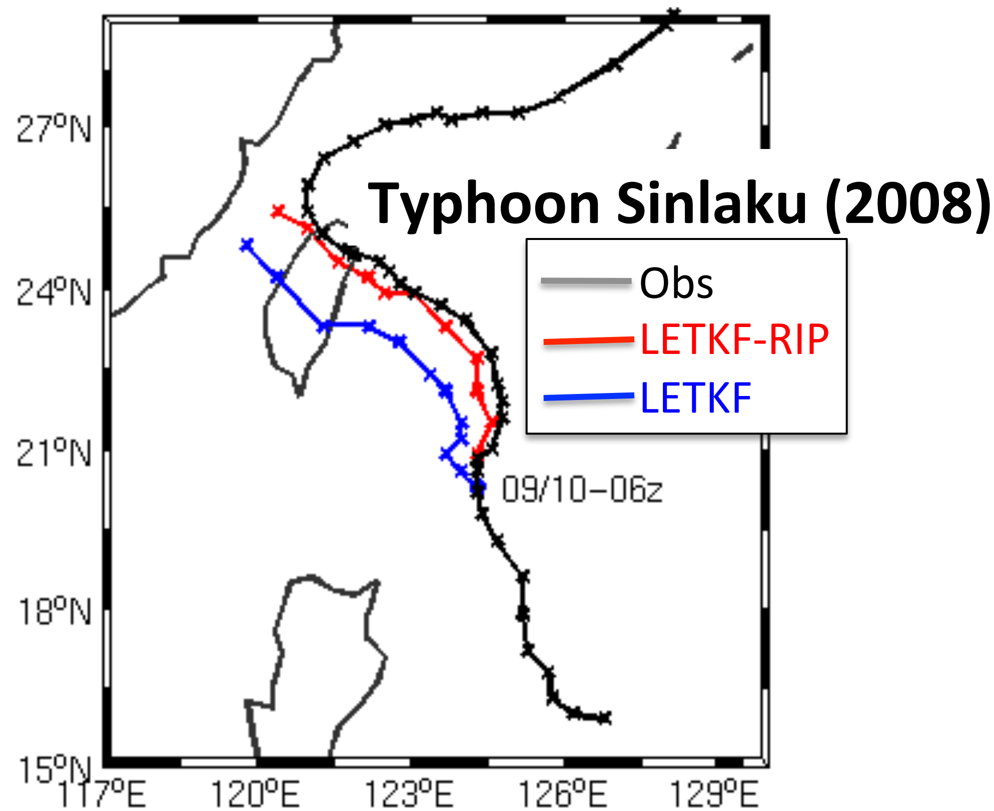
- RIP adapts to using an observation N-times by dividing the spread by N: RIP converges to the regular optimal KF solution.
- The spin-up is faster and the analysis update is "softer" (in small steps) rather than in large steps.

# 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

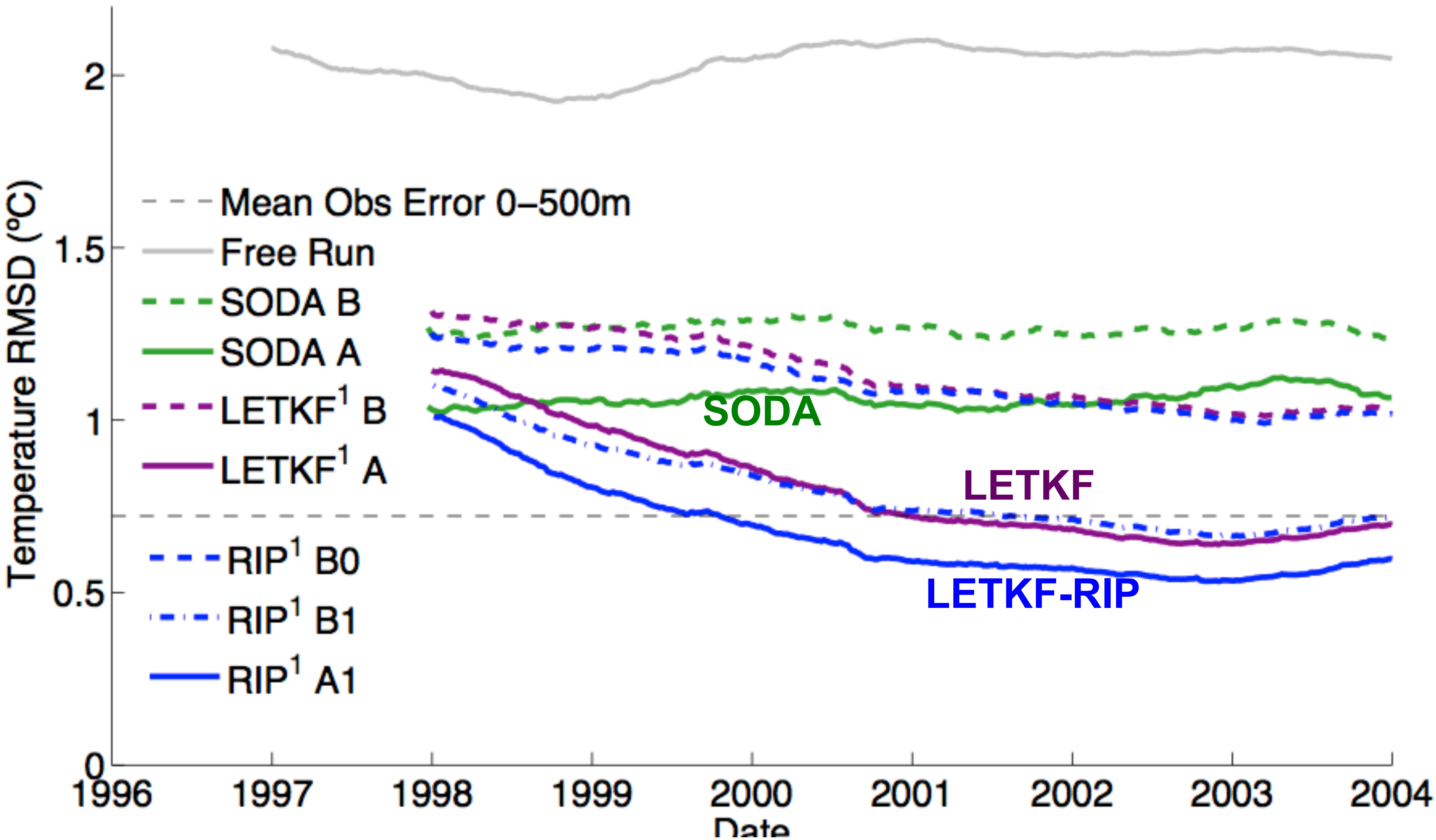
Advisors: E Kalnay, J Carton, K Ide, T Miyoshi, G Chepurin

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

RMSD (°C) (All vertical levels)

# 7 years of Ocean Reanalysis Temperature

B: background
A: analysis

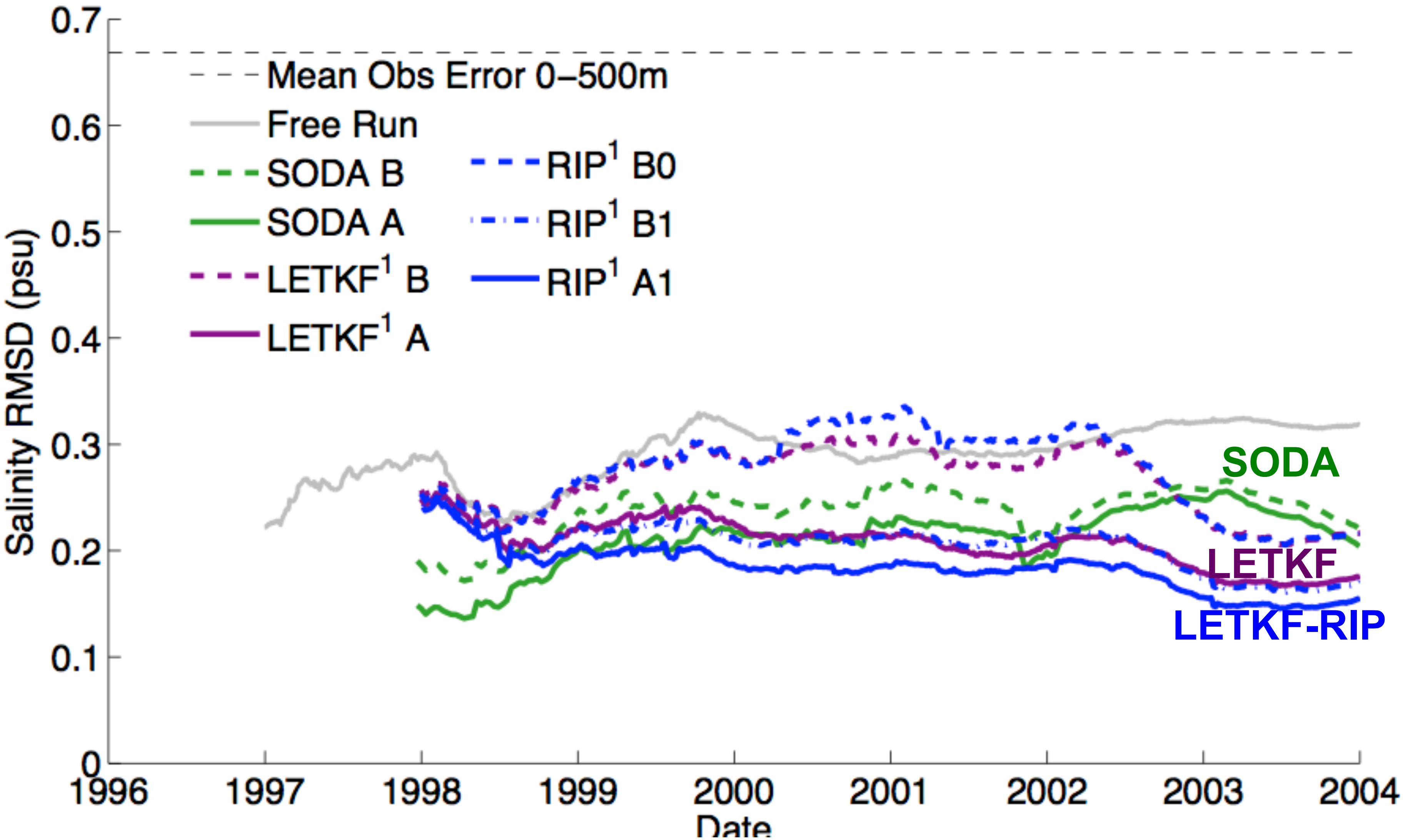



Global RMS(O-F) of Temperature (°C),
SODA, LETKF, RIP

# 7 years of Ocean Reanalysis Salinity

RMSD (psu) (All vertical levels)

B: background
A: analysis




Global RMS(O-F) of Salinity (psu),
SODA, LETKF, RIP

# Summary for LETKF-RIP (or QOL)

- Kalman Filter is optimal for a linear, perfect model.
- During spin-up, or when the ensemble perturbations grow nonlinearly, EnKF is not optimal, since it does not extract enough information from the observations.
- The LETKF "no-cost" smoother (or, equivalently, the 4D-EnSRF) allows LETKF-RIP to use the observations more than once, and thus extract much more information.
- This shortens the spin-up and produces more accurate forecasts with the same observations.
- For linear models RIP converges to the same optimal KF solution but with spread reduced by ~ $\sqrt{N}$
- For long windows and nonlinear perturbations, RIP advances in smaller steps and approaches the true attractor more "softly".

# (2) 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!

# Transform precipitation $y$ into a Gaussian $y_{transf}$

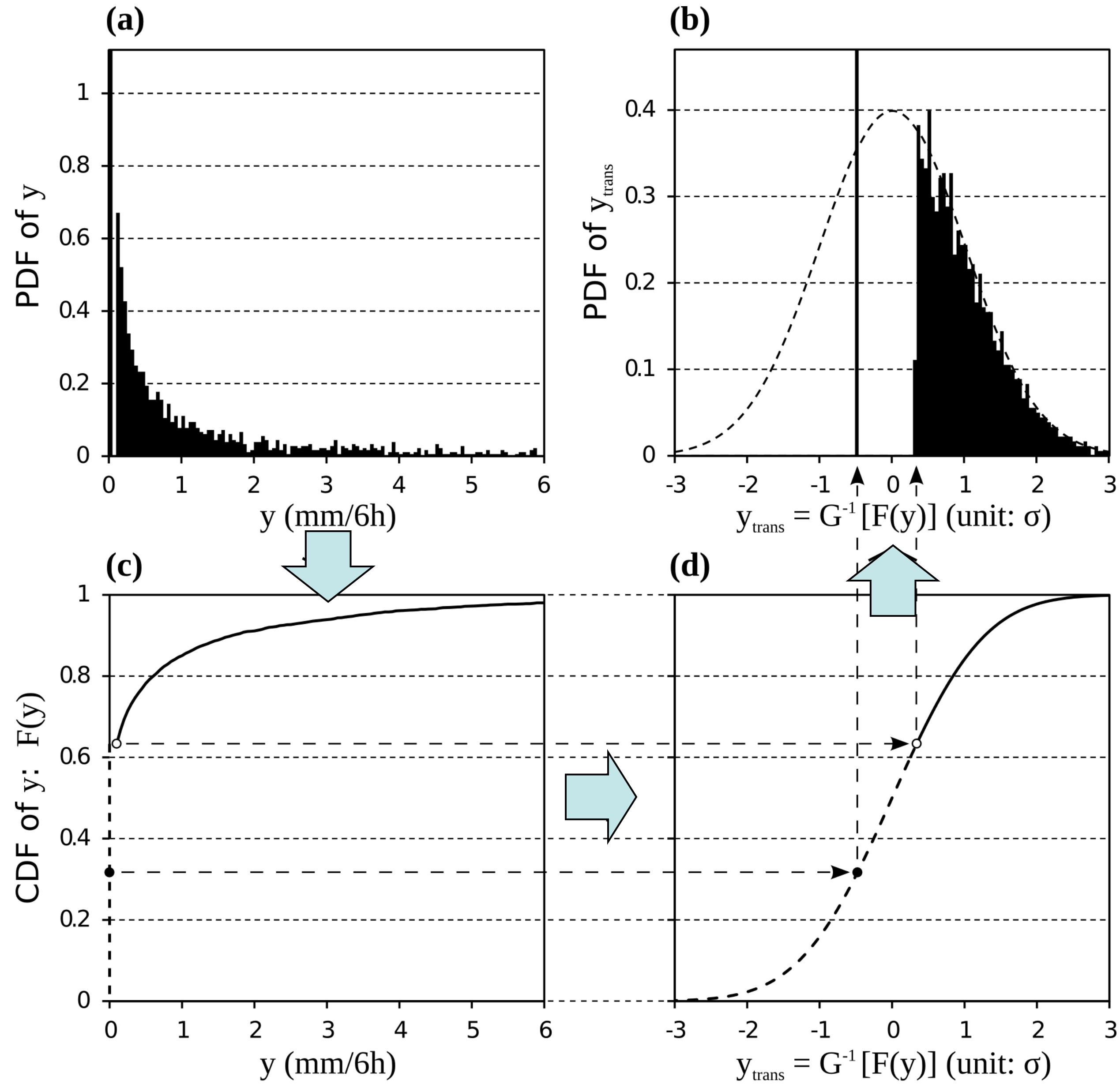
Start with pdf of $y$=rain at every grid point.

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

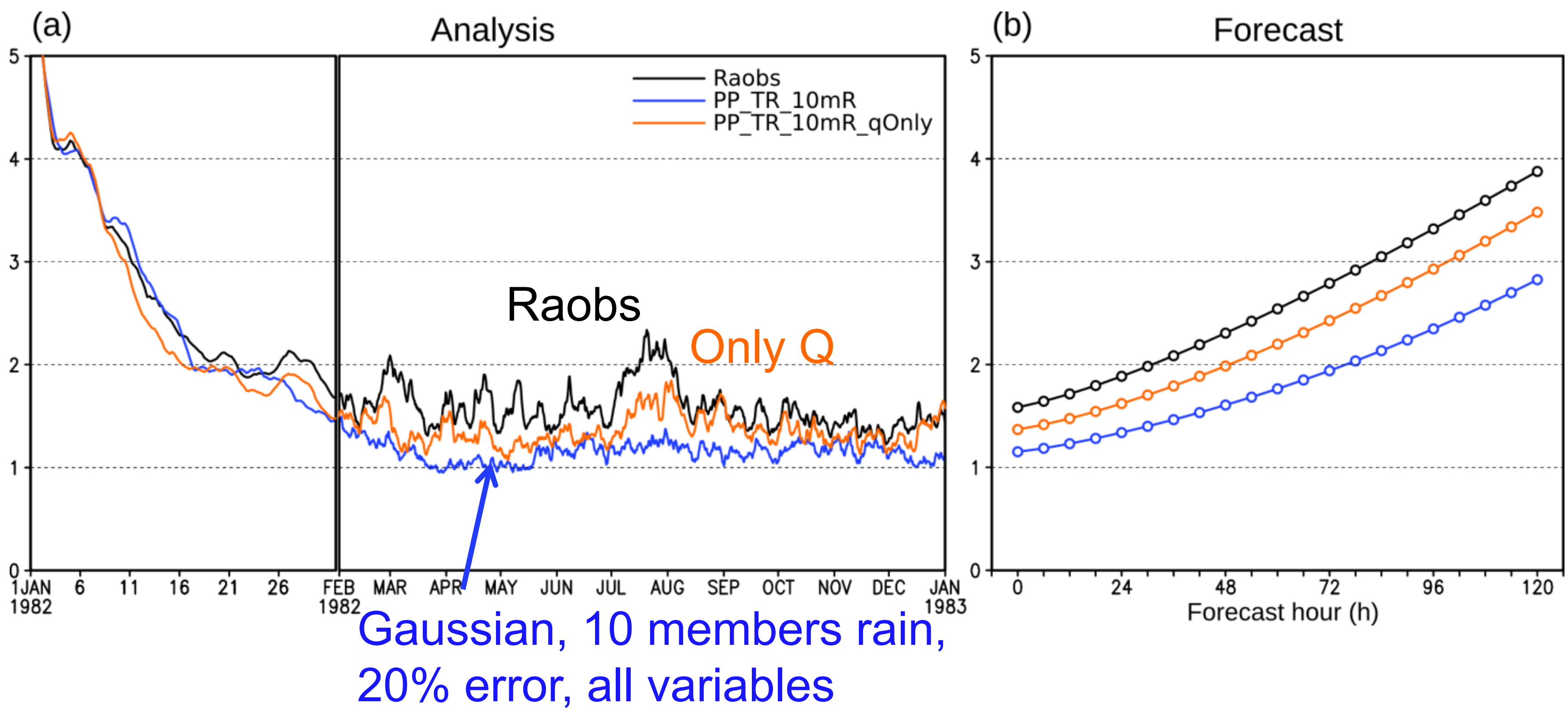
We assign all "no rain" to the **median** of the no rain CDF.

We found this works as well as more complicated procedures.

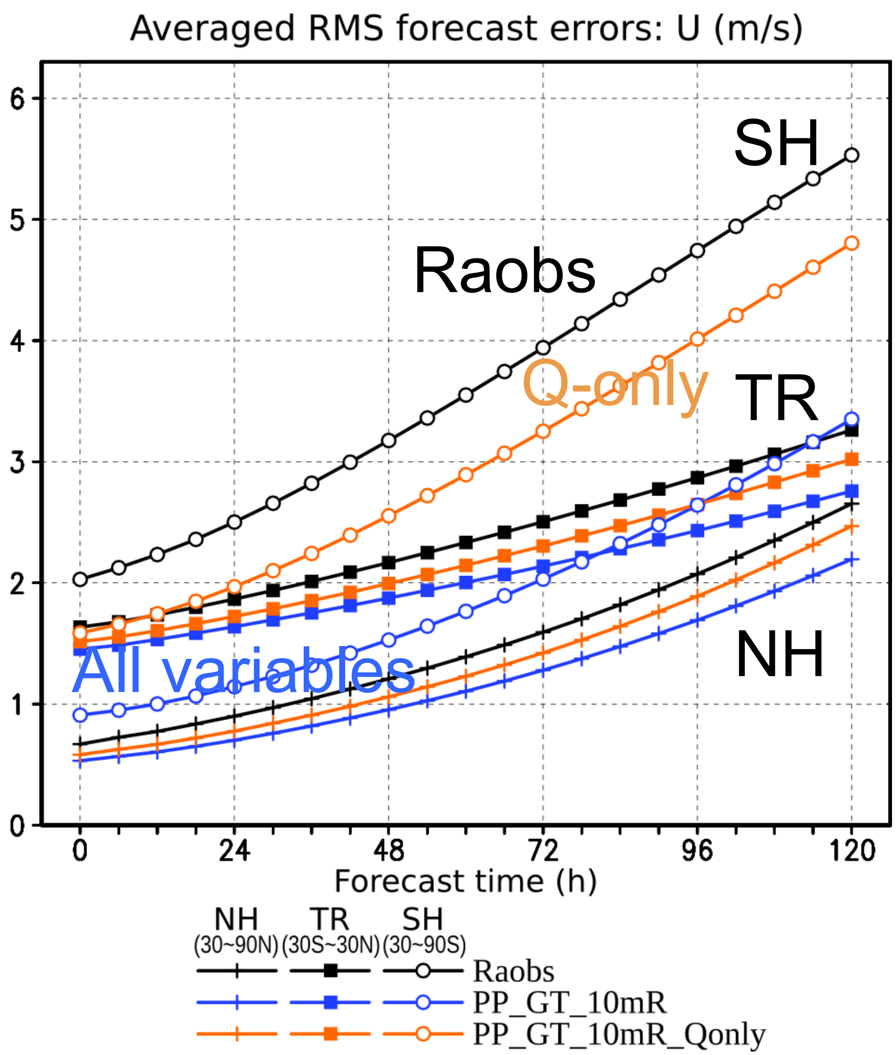
It allows to assimilate both rain and no rain.




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

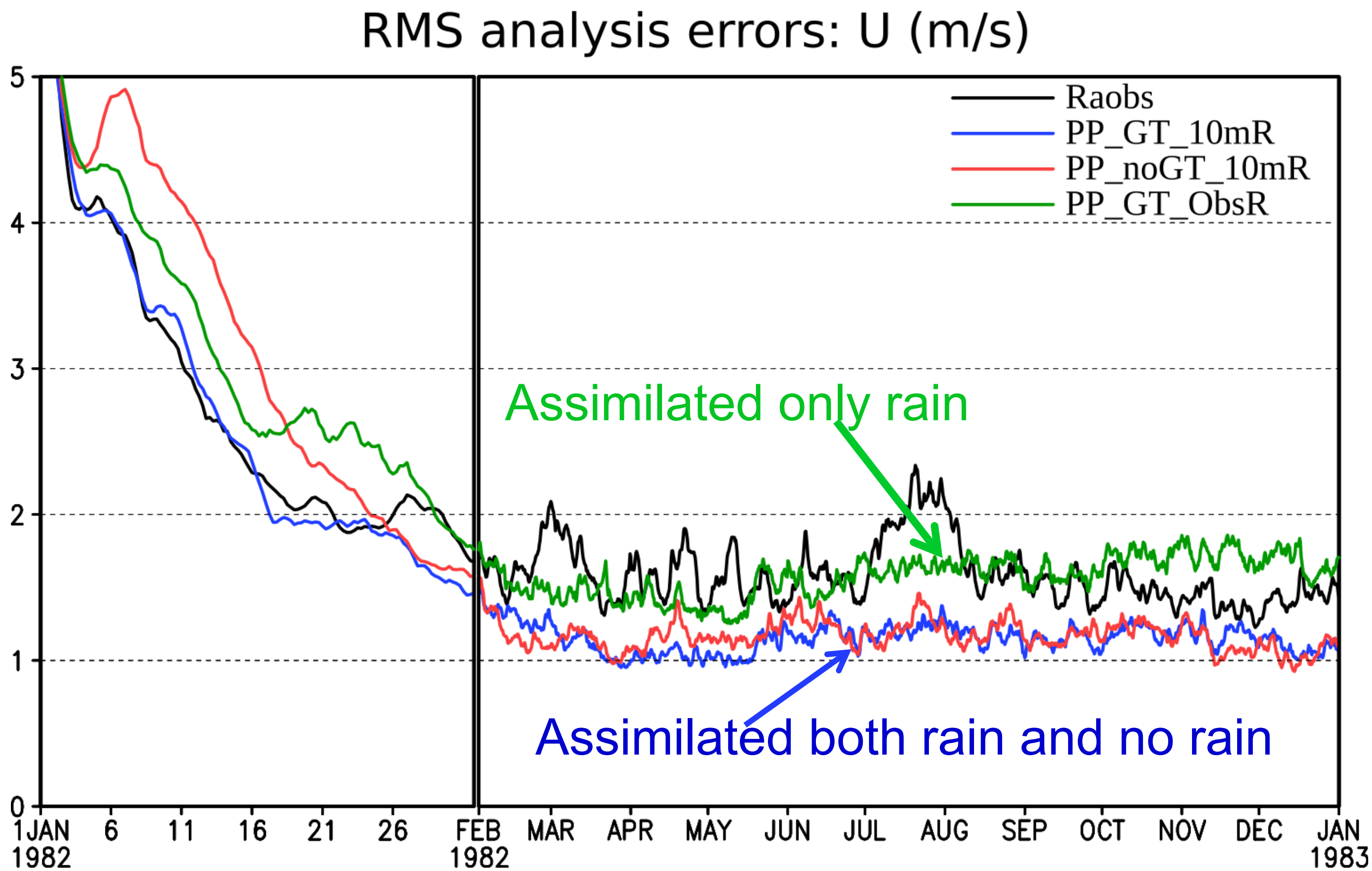



- Main result: with at least 10 ensemble members raining in order to assimilate an obs, updating all variables (including vorticity), with Gaussian transform, and rather accurate observations (20% errors), the analyses and forecasts are much improved!
- Updating only Q is much less effective.
- **The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

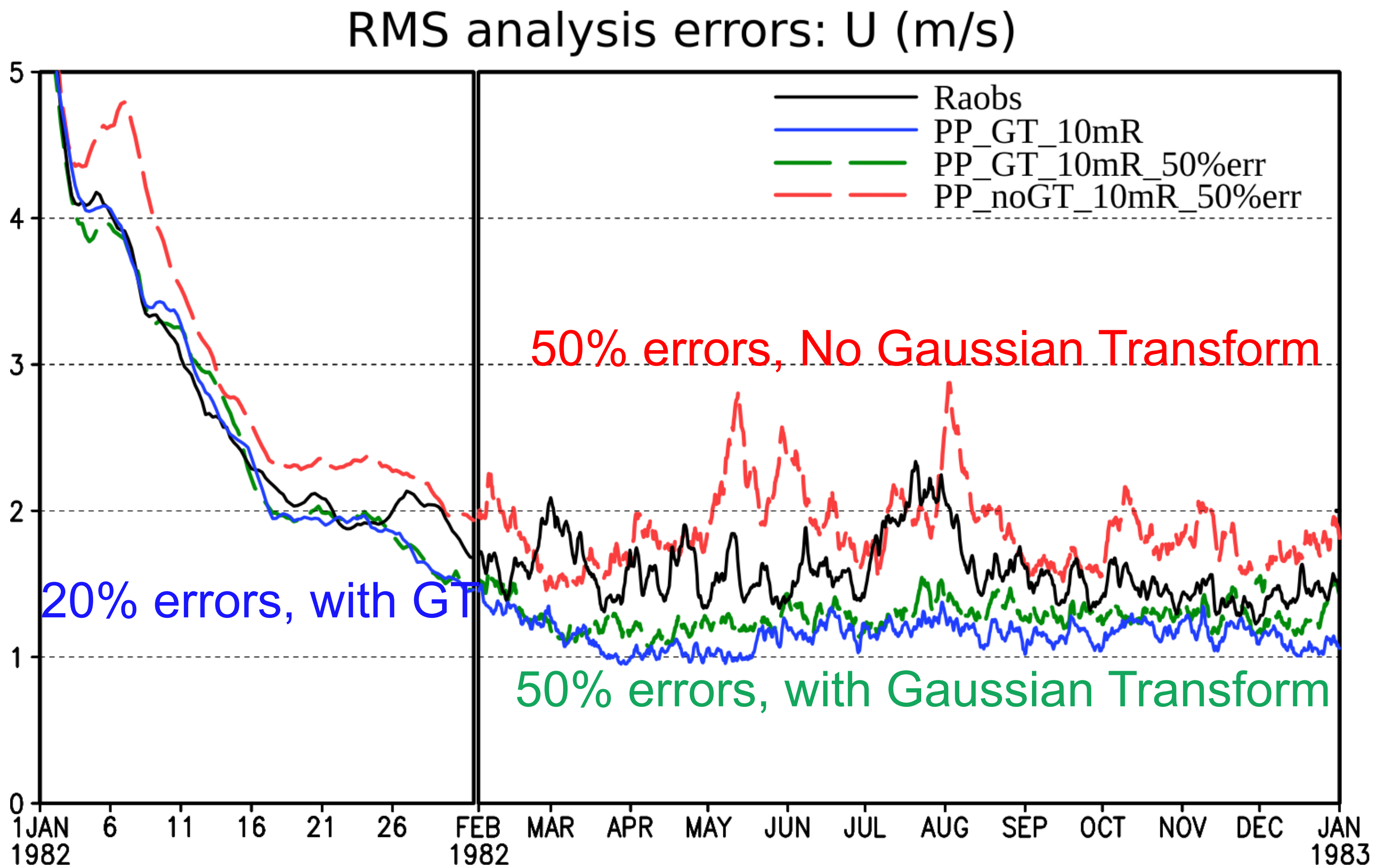



One year of 5-day forecasts

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




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

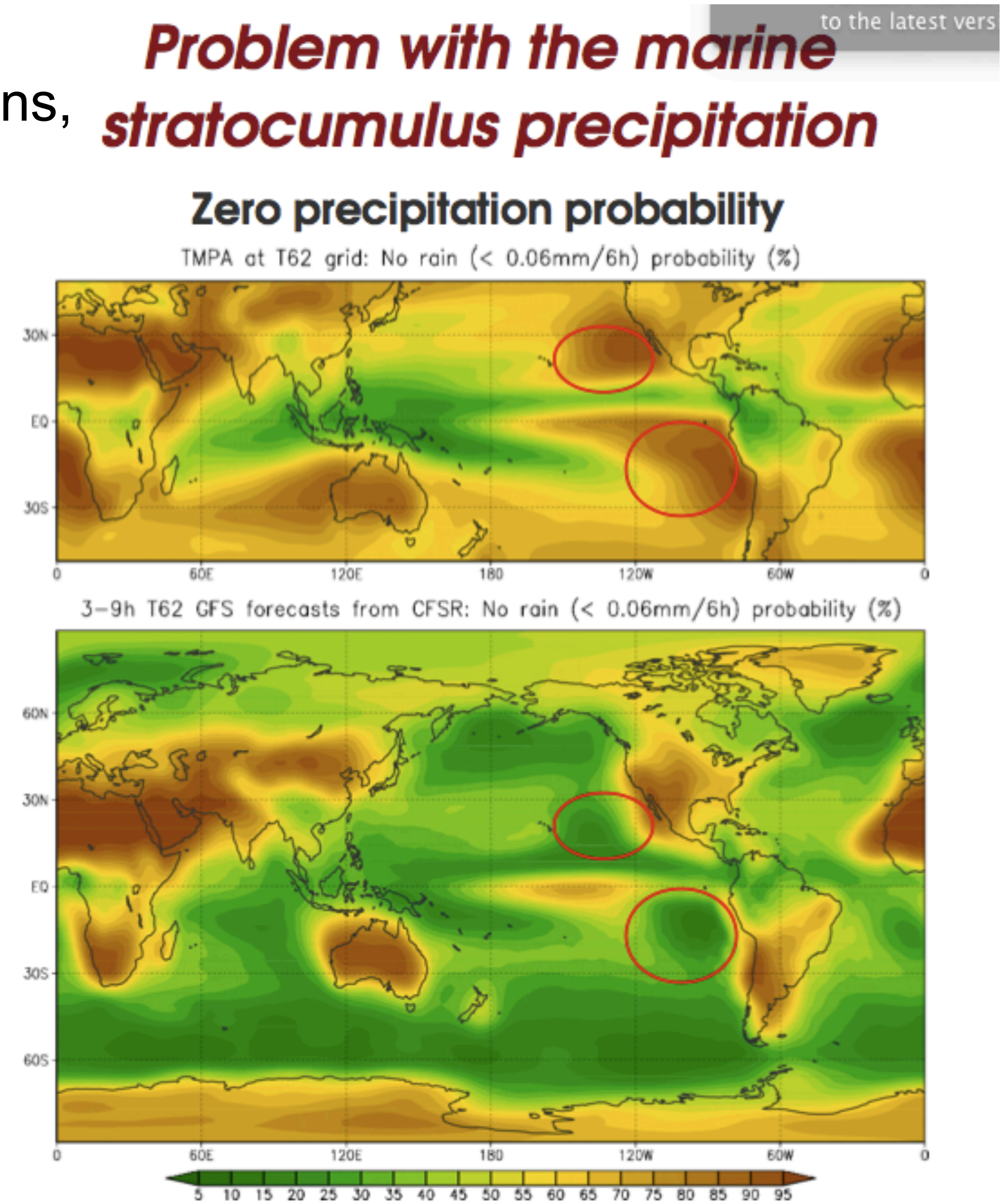



The impact of the Gaussian Transform is important with larger observation errors (50% rather than 20%). The impact of GT50% is almost as good as GT20%.

Real observations, model errors

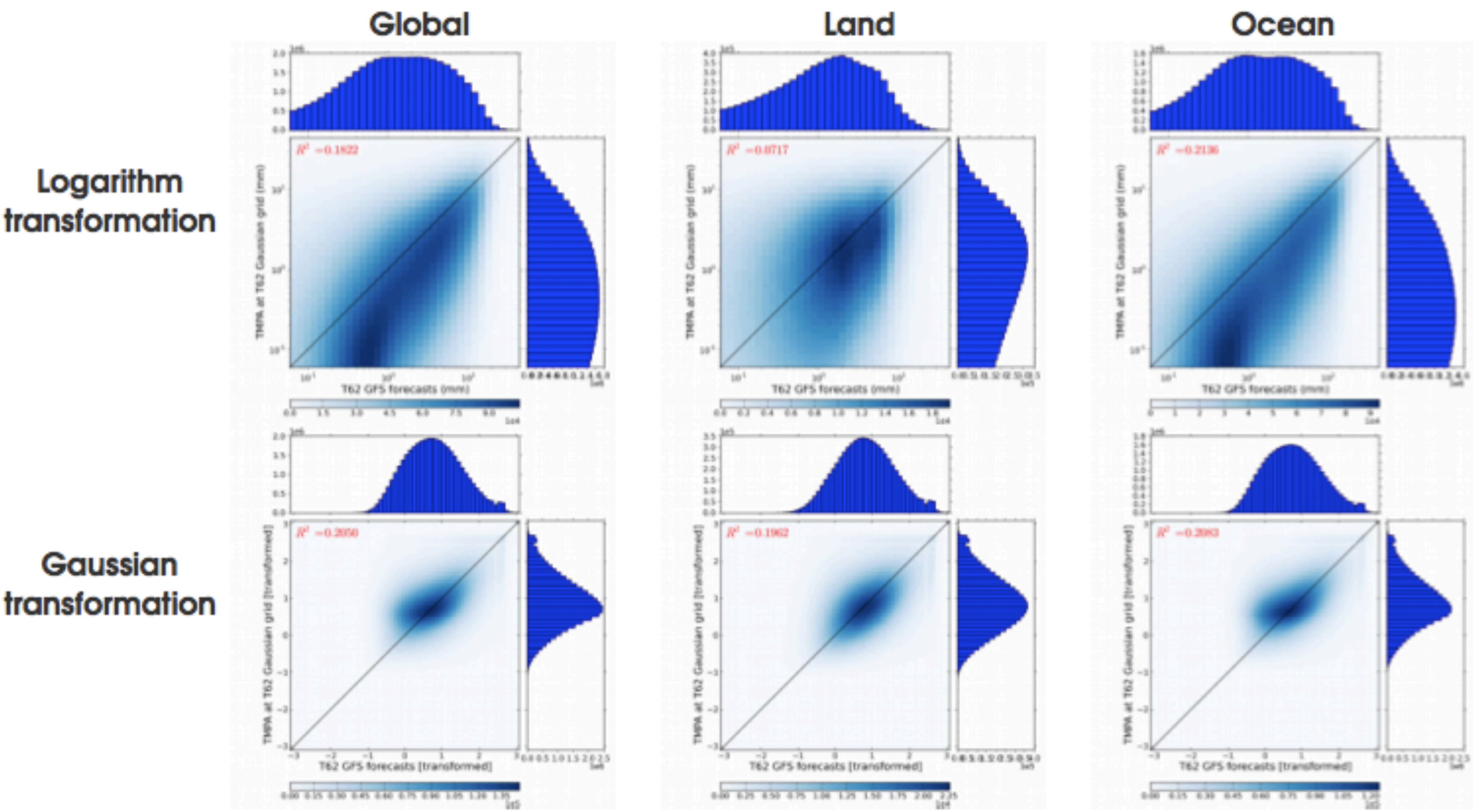
TMPA

GFS T62



# TMPA (TRMM+) statistics vs GFS T62

## *Logarithm transformation v.s. Gaussian transformation*



# Summary for assimilation of precipitation

- The model remembers potential vorticity (dynamics), does not remember moisture changes, or even temperature.
- EnKF has a better chance to assimilate potential vorticity by giving higher weights to ensemble members with right precip.
- EnKF also does not require 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 in a perfect model.
- We are attempting to assimilate TMPA precip in the GFS, but can only afford T62!
- Will try to assimilate only synoptic scale rain.
- Following Philippe Lopez we will also try assimilating observed precip over North America.

# Ensemble Forecast Sensitivity to Observations (EFSO) and Proactive QC

Eugenia Kalnay[1], Yoichiro Ota[2,3], **Daisuke Hotta**[1,2], Takemasa Miyoshi[4,1]

(1) University of Maryland

(2) Japan Meteorological Agency

(3) National Centers for Environmental Prediction

(4) RIKEN, Kobe, Japan

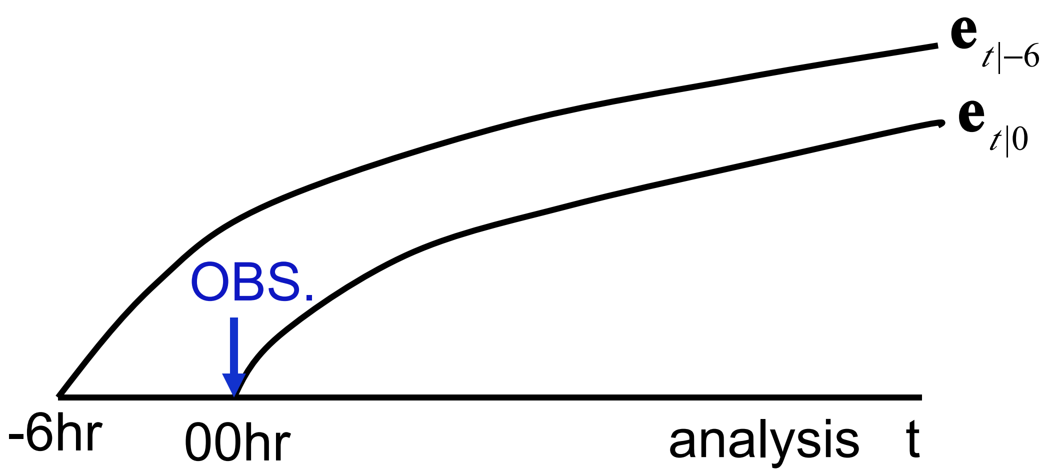
**Acknowledgements:** NESDIS JPSS for support, and JCSDA for access to S4, Dr. Jung for advice

# 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, Liu: Tellus, 2012)




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

(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:

$$(\overline{\mathbf{x}}_0^a - \overline{\mathbf{x}}_{0|-6}^b) = \mathbf{K}(\mathbf{y} - H(\mathbf{x}_{0|-6}^b))$$

- Observation impact on the reduction of forecast error:

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

# 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 and Zhu, 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

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

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

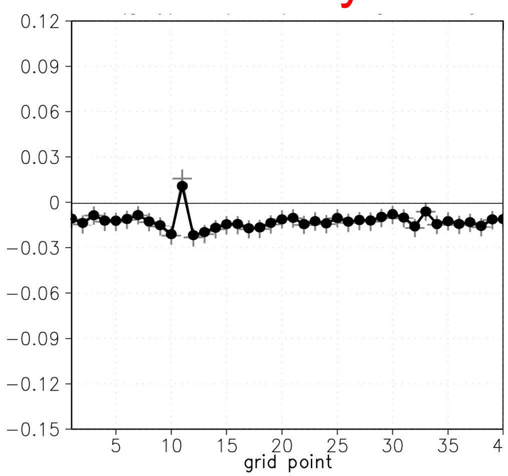
$$\boxed{\begin{aligned} \Delta \mathbf{e}^2 &= \left[ \mathbf{MK}(\mathbf{y} - H(\mathbf{x}^b_{0|-6})) \right]^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\ &= \left[ (\mathbf{y} - H(\mathbf{x}^b_{0|-6}) \right]^T \mathbf{R}^{-1} \mathbf{Y}^a_0 \mathbf{X}^{fT}_{t|0} (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})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

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

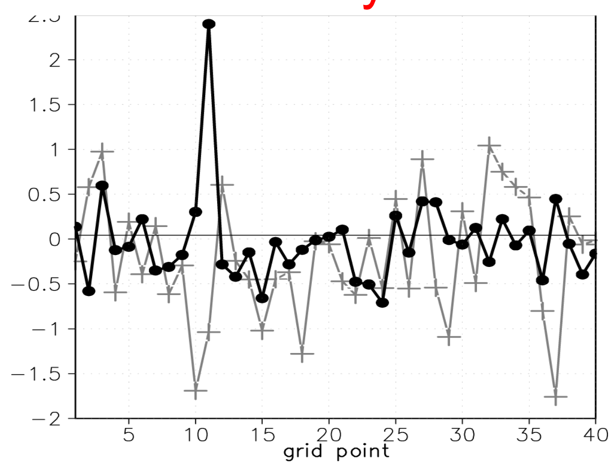
# Tested ability to detect a poor quality ob impact on the forecast in the Lorenz 40 variable model

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

1 day



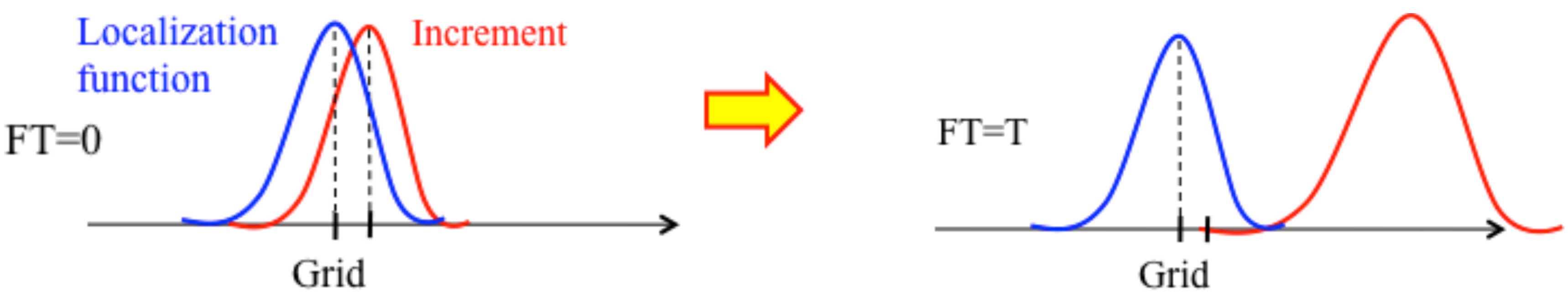

10 days




✓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: <u>an approximation</u>

# Ota et al. 2013: Applied EFSO to NCEP GFS/ EnSRF using all operational observations.

## Preliminary criteria used to identify regional 24hr "forecast failures"

- Divide the globe into $30^o$x$30^o$ 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 failures. In every case, the forecast improved without the "bad observations".

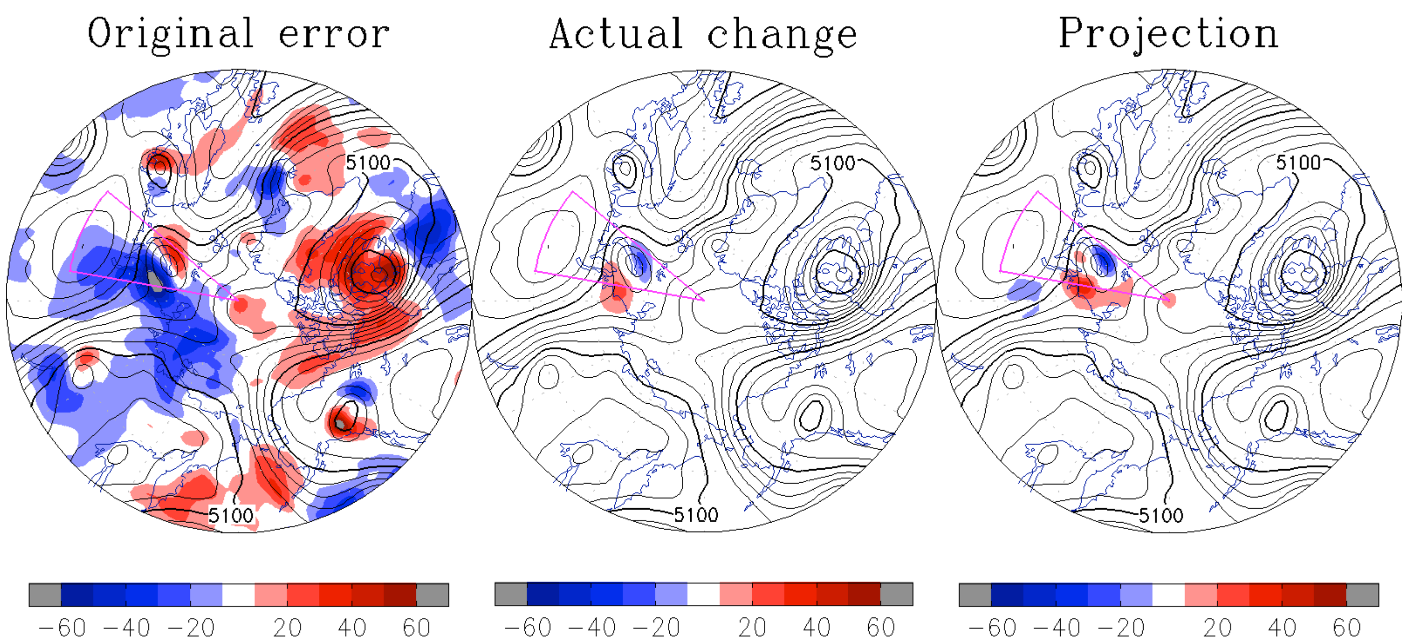
# 24-hr forecast error correction (Ota et al. 2013)

- identified 7 cases of large $30^o$x$30^o$ regional errors,
- rerun the forecasts denying bad obs.
- the forecast errors were substantially reduced

| 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% |

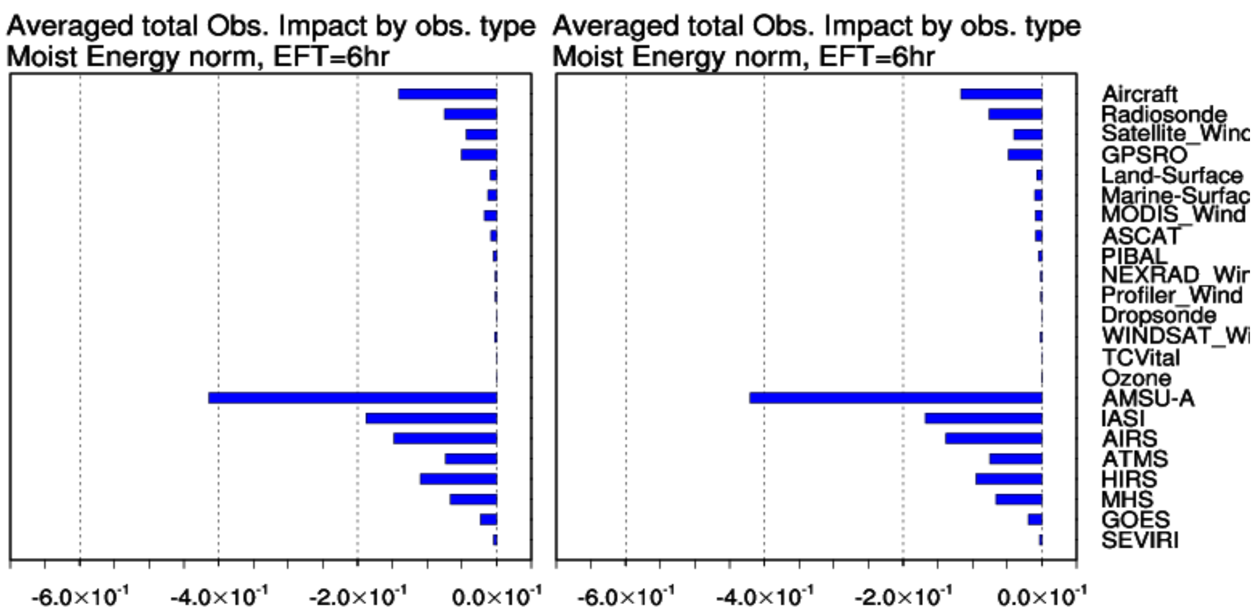
MODIS

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




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

# **Daisuke Hotta**: Did 18 days of EFSO but using the LETKF Hybrid (not the EnSRF), and 6 hr forecasts, not 24 hr forecasts.



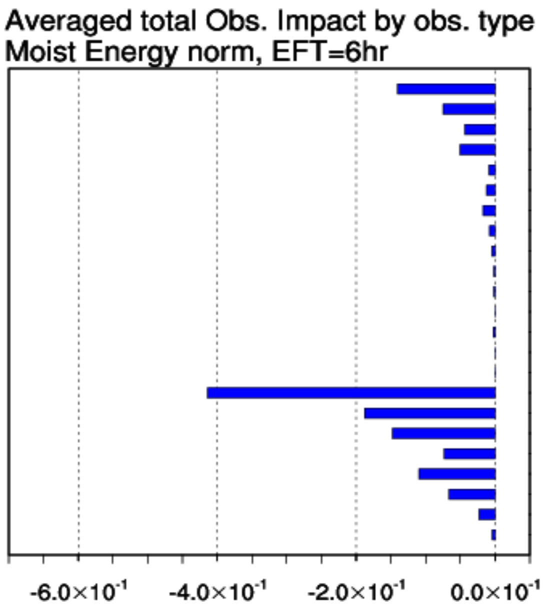

Evaluation time: 6 hours
**Verified against LETKF analysis**

Evaluation time: 6 hours
**Verified against hybrid GSI analysis**

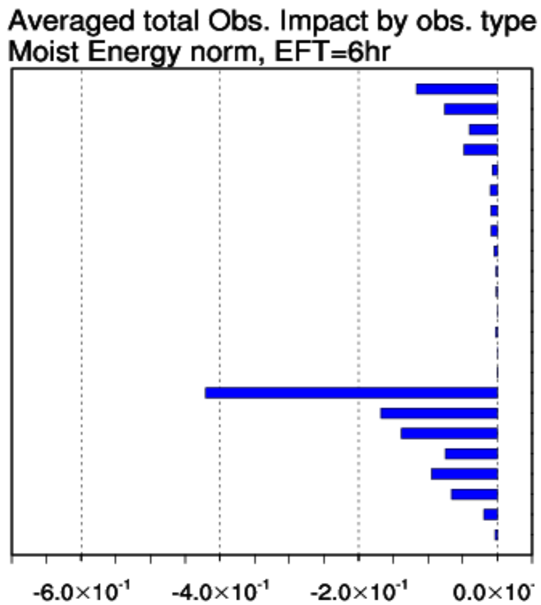
**Time averaged EFSO's are rather insensitive to the verifying analysis!**

# Average total observation impact:

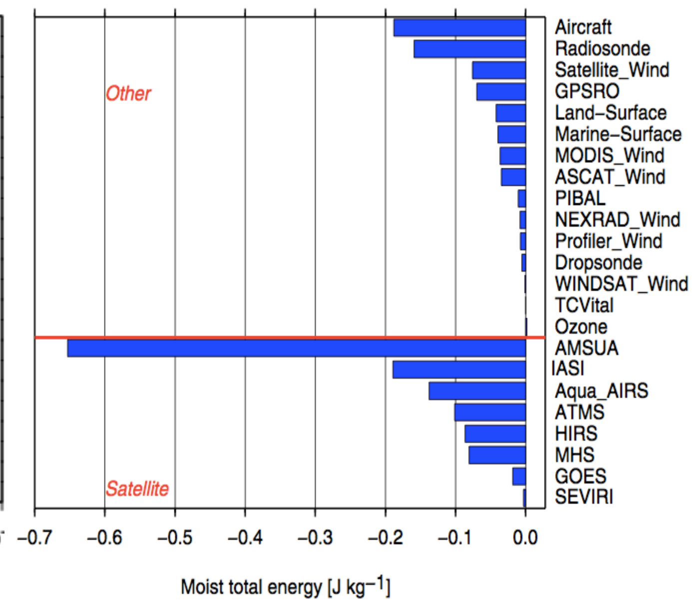
# Comparison with Ota et al. (2013), 31 days, using the EnSRF verification after 24hrs, not 6 hrs.




Evaluation time: **6 hours**
**Verified against**
**LETKF analysis**

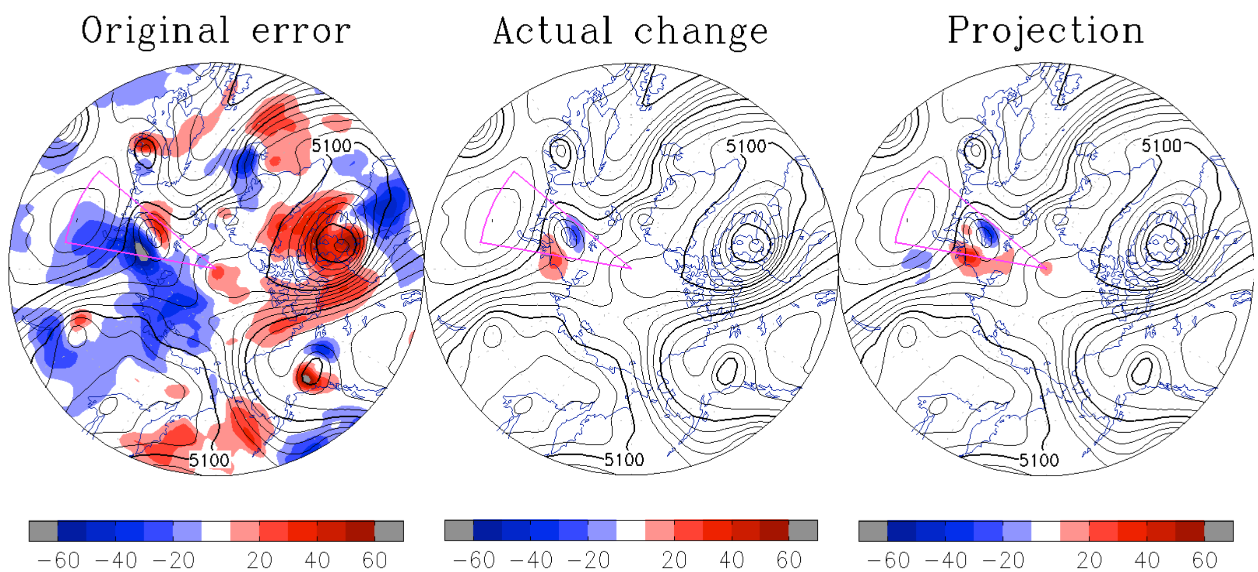


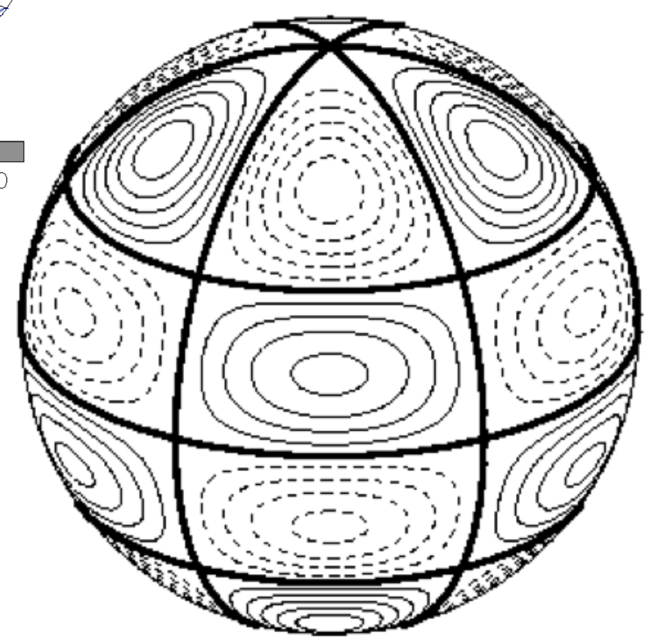

Evaluation time: **6 hours**
**Verified against**
**hybrid GSI analysis**




Evaluation time: **24 hours**
**Verified against**
**EnSRF analysis**
**from Ota et.al (2013)**

# In the future, we will define regions of uniform areas using spherical harmonics (e.g., n=6, m=3)



# EFSO and Proactive QC: Summary

- Ensemble Forecast Sensitivity to Observations (**EFSO**, Kalnay et al., 2012, Tellus) is a simpler and more accurate formulation than Liu and Kalnay (2008, QJRMS).
- Ota et al., 2013, Tellus tested EFSO with the NCEP EnSRF-GFS operational system using all operational observations.
- **EFSO was used to identify "bad observations" with large negative regional impacts after 24hr.**
- **Hotta has shown that we can use EFSO with 6hr forecasts from a hybrid, and that it is not too sensitive to the choice of verification.**
- **"Proactive QC": repeat the 6hr data assimilation without the identified bad obs, and**
- **Save the "bad obs" with metadata from EFSO and provide them to the algorithm developers.**

# Applications of ensemble singular vectors in the LETKF framework

Shu-Chih Yang and Eugenia Kalnay
with thanks to
T. Enomoto

# Ensemble Singular Vectors

Define the vector of initial (I) and final (F) perturbations:

$$\mathbf{X}_{t-\Delta t}^{I} = \left[\delta\mathbf{x}_{1,t-\Delta t}, \cdots, \delta\mathbf{x}_{i,t-\Delta t}, \cdots, \delta\mathbf{x}_{K,t-\Delta t}\right]; \quad \mathbf{X}_{t}^{F} = \left[\delta\mathbf{x}_{1,t}^{F}, \cdots, \delta\mathbf{x}_{i,t}^{F}, \cdots, \delta\mathbf{x}_{K,t}^{F}\right]$$

Find the linear combination of initial perturbations that will grow fastest given a optimization time period

Initial ES: $\delta\mathbf{x}_{t-\Delta t}^{I} = \mathbf{X}_{t-\Delta t}^{I}\mathbf{p}$

Final ES: $\delta\mathbf{x}_{t}^{F} = \mathbf{X}_{t}^{F}\mathbf{p}$

By defining the initial and final perturbation norms ($\mathbf{C}_I$ and $\mathbf{C}_F$), we can solve $\mathbf{p}$ (Enomoto et al. 2006).

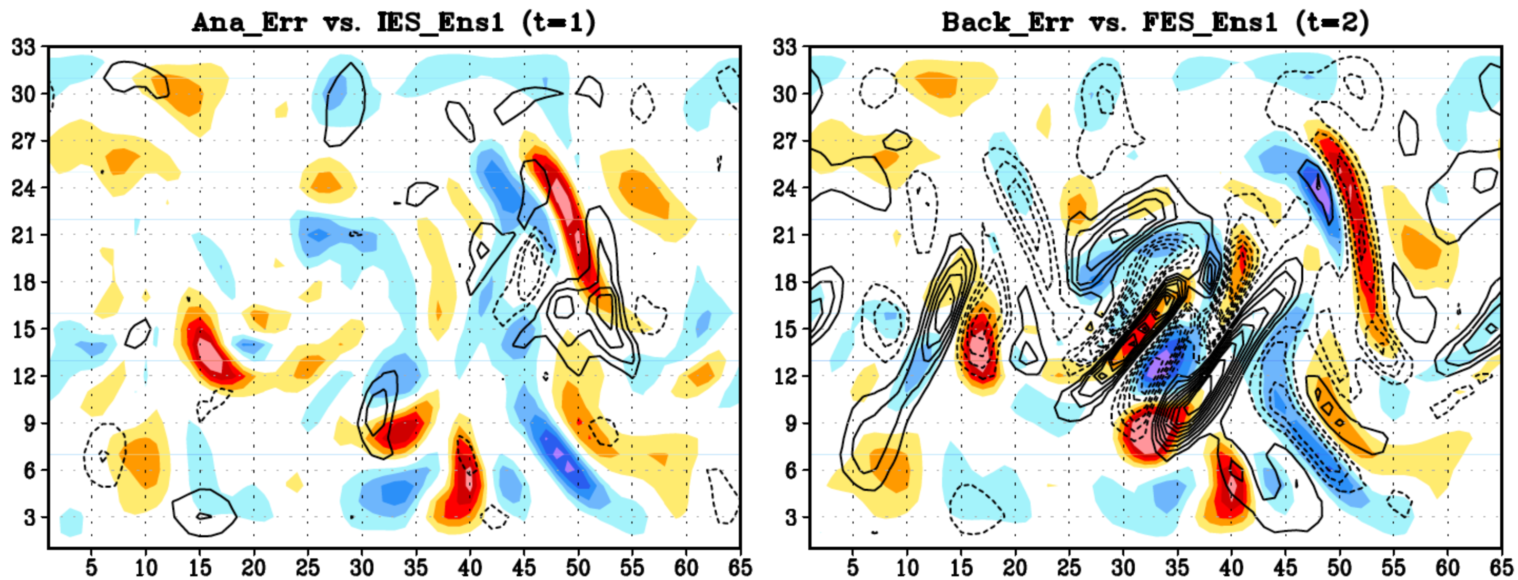
$$\boxed{\left(\mathbf{X}_{t-\Delta t}^{I}{}^{T}\mathbf{C}_I\,\mathbf{X}_{t-\Delta t}^{I}\right)^{-1}\left(\mathbf{X}_{t}^{F}{}^{T}\mathbf{C}_F\,\mathbf{X}_{t}^{F}\right)\mathbf{p} = \lambda\,\mathbf{p}}$$

We can find $K$ sets of IEI and FES with $\left(\lambda^{i}, \mathbf{p}^{i} \quad i = 1, \cdots, K\right)$

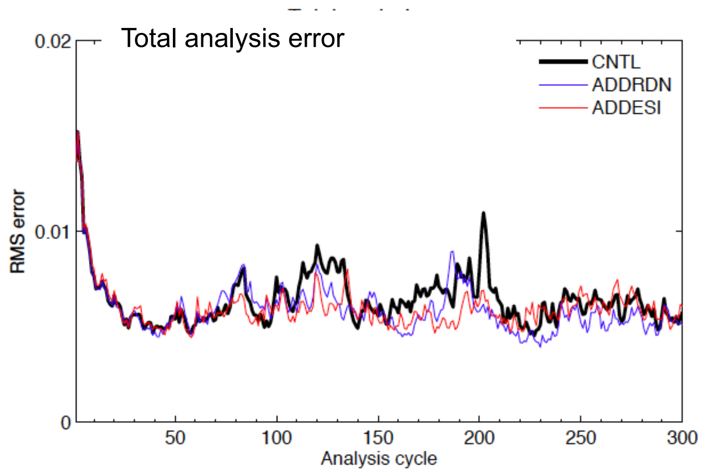
# Ensemble sensitivity (ES) in a Quasi-geostrophic model

$$X^I_{t-\Delta t} = X^a_{t-\Delta t}$$ (LETKF Ana. Ens); $$X^F_t = X^b_t$$ (LETKF Back. Ens )

**Ensemble Sensitivity with an 12-hr interval**

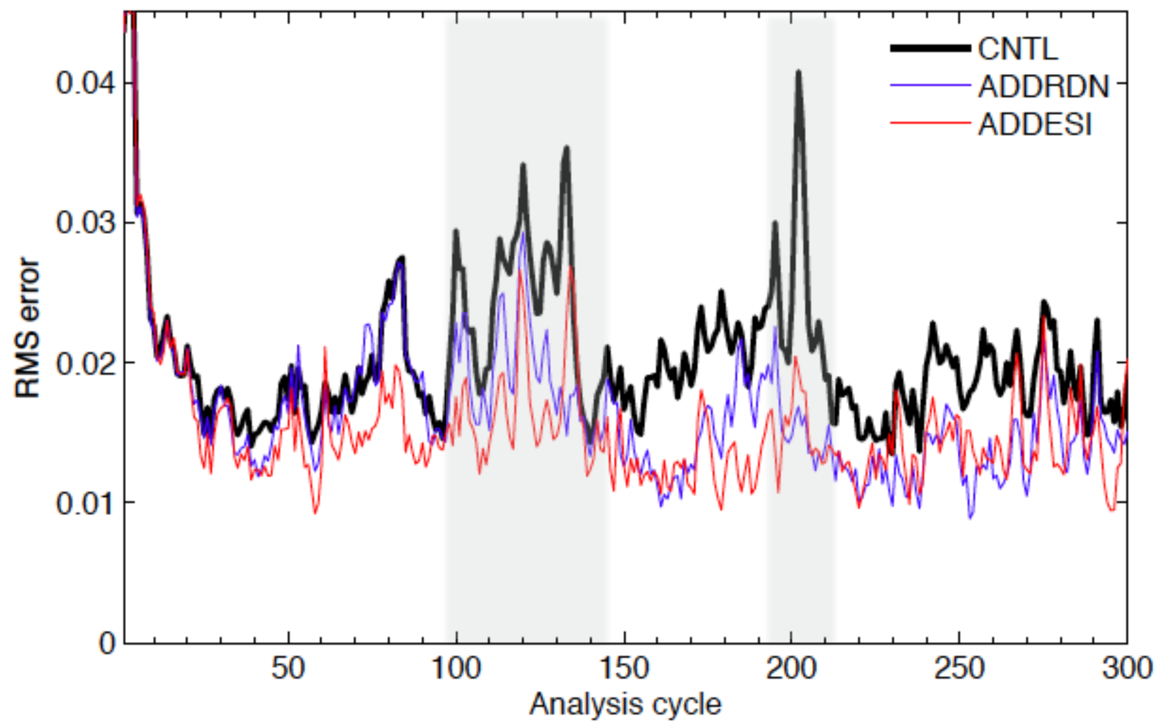



The fast growing perturbation (contours) is very closely related to the background errors (color). The IES (an initial Singular Vector) is NOT related to the initial errors.



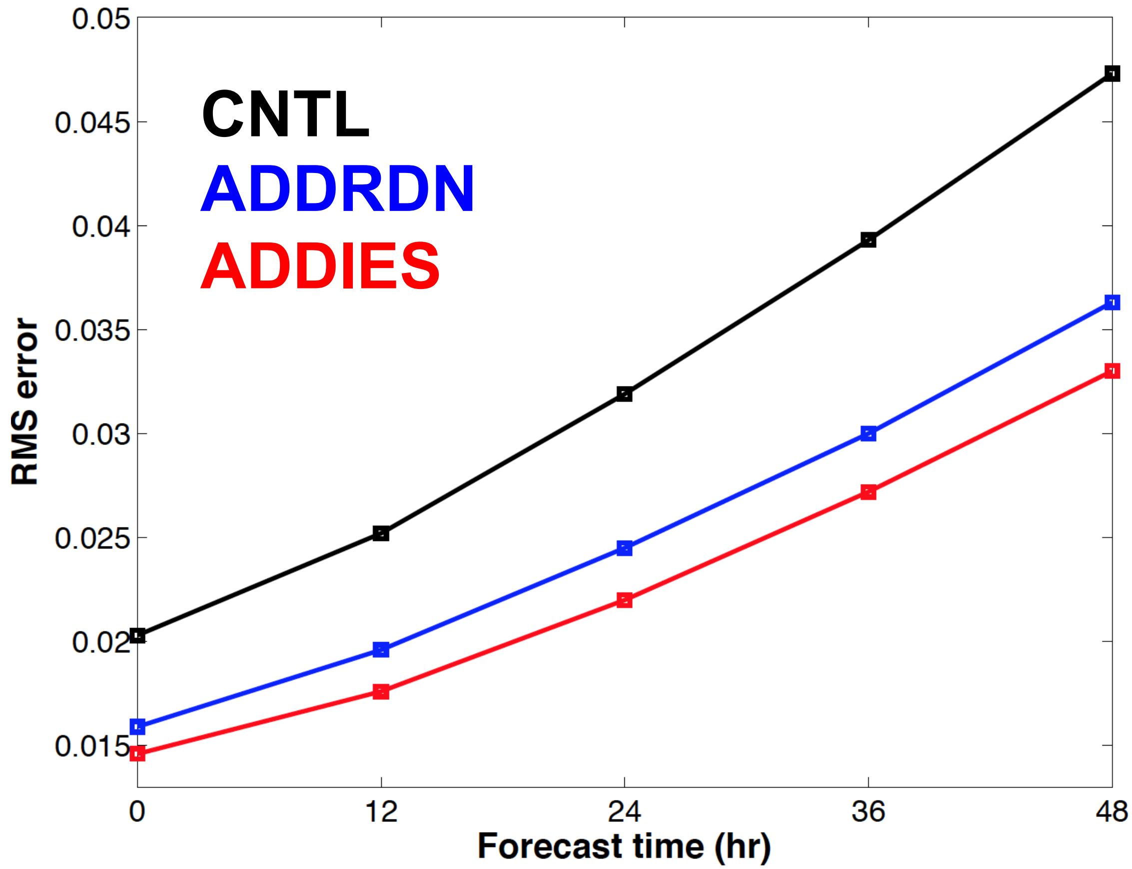

Used IES as additive inflation

Large analysis error




ADDIES is particularly effective in correcting fast growing errors!!

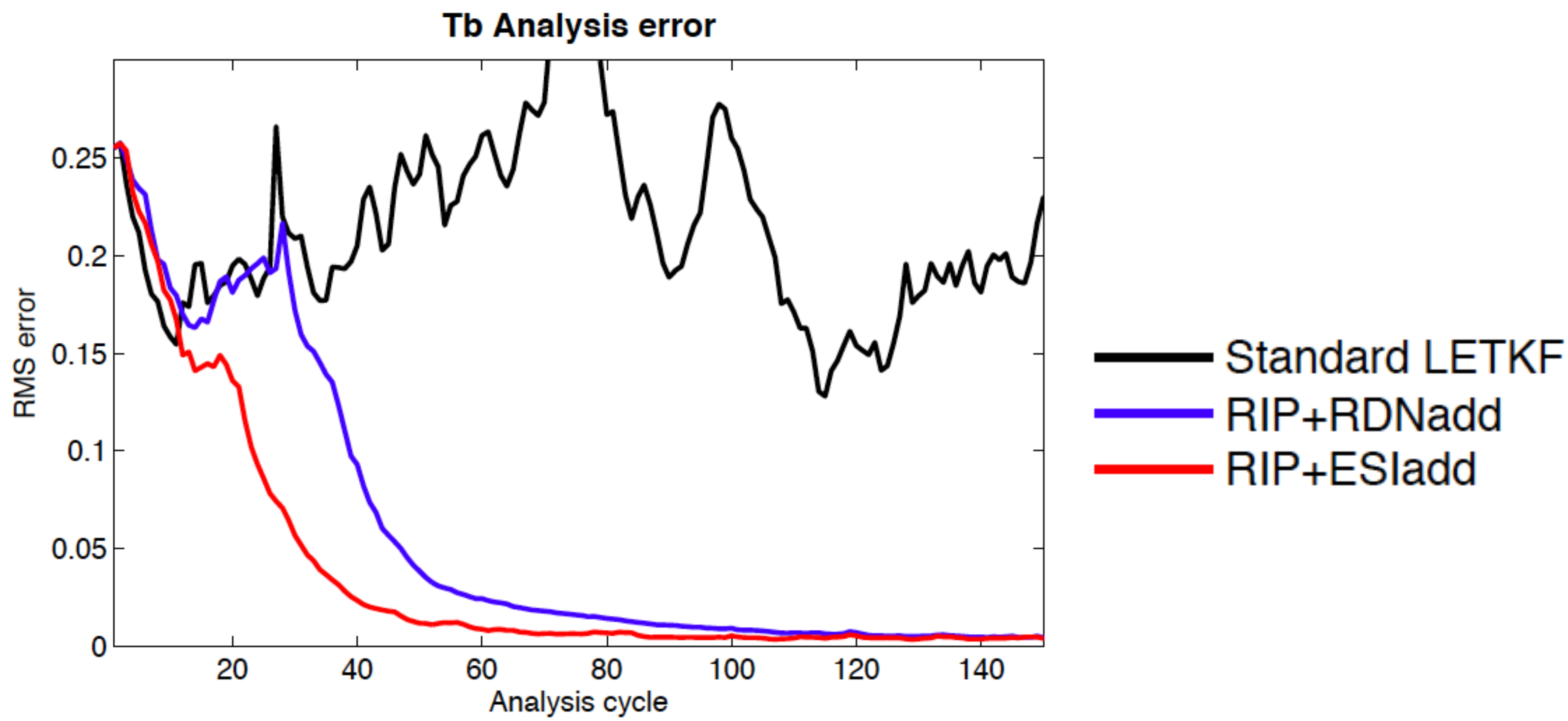
# Growing errors




- In CNTL, the incompletely removed growing errors **amplify** at later forecast time.
- ADDIES successfully removes growing errors!

# Correcting fast growing errors with LETKF-RIP

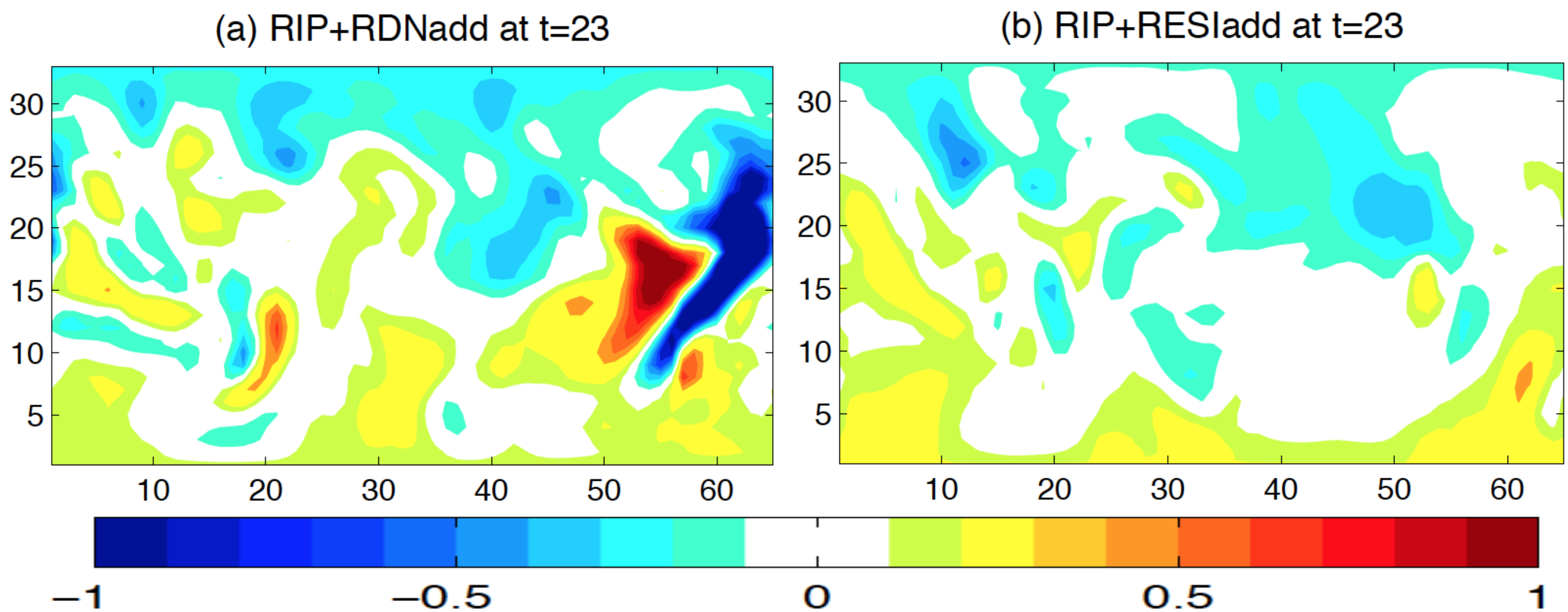
IESs are used as the additive inflations for refreshing the smoothed analysis during RIP iterations.




IES is more effective than the random perturbations and further accelerates the LETKF's spin-up!

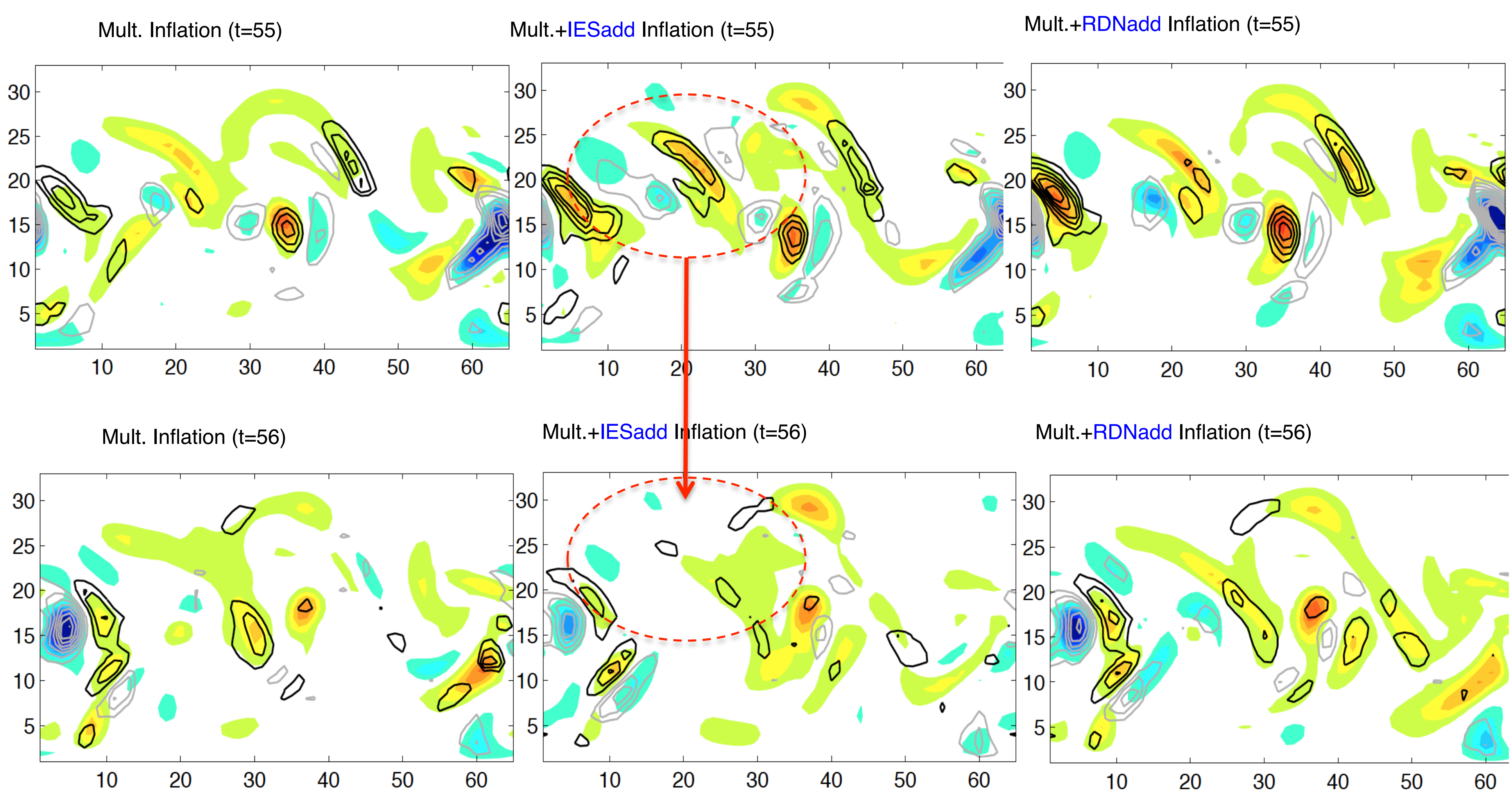
# Correcting fast growing errors with LETKF-RIP

## Analysis error during LETKF's spin-up




Large errors can be quickly removed in the RIP+ ESIadd analysis!

# Background error (color) vs. analysis increment (contour)



# Simultaneous data assimilation of $CO_2$ and meteorological variables within LETKF coupled with NCAR CAM model

[*]**Ji-Sun Kang**, [*]Eugenia Kalnay, [+]Junjie Liu, [#]Inez Fung, and [*]Takemasa Miyoshi

[*]University of Maryland, College Park,MD

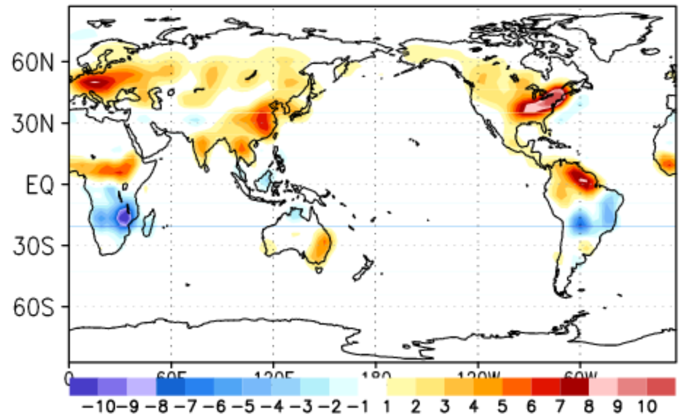
[+]NASA/JPL, Pasadena, CA

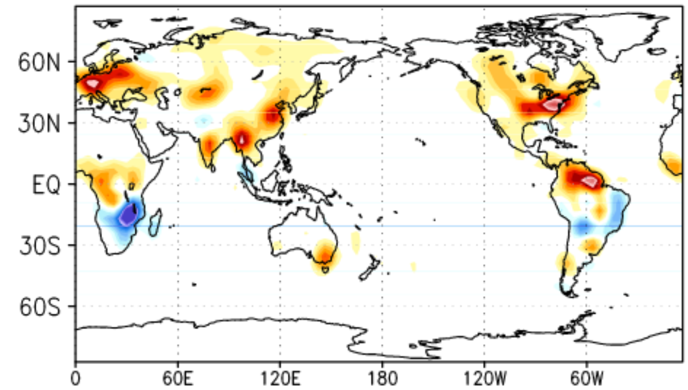
[#] University of California, Berkeley, CA

# Results

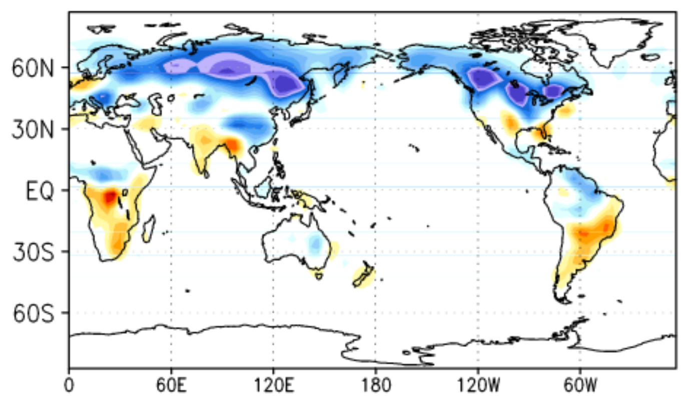
▼ True CF ▼ ▼ Analysis of CF ▼

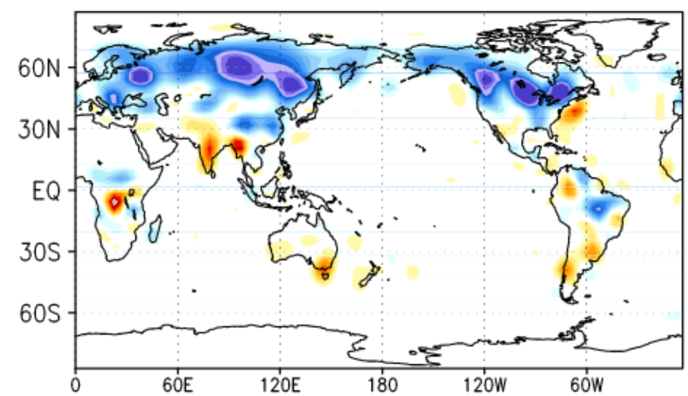
00Z01APR ▶
After three months of DA



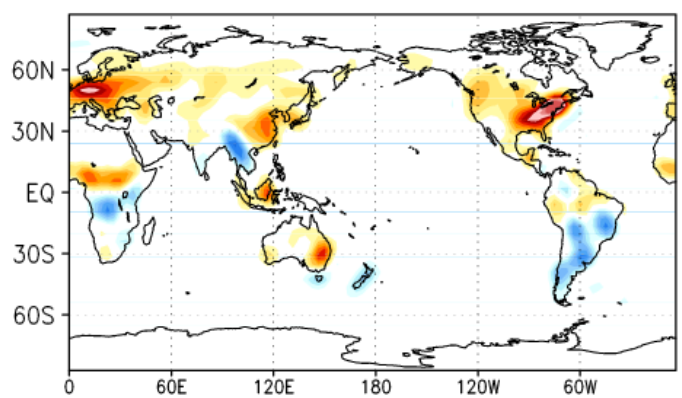

00Z01AUG ▶
After seven months of DA

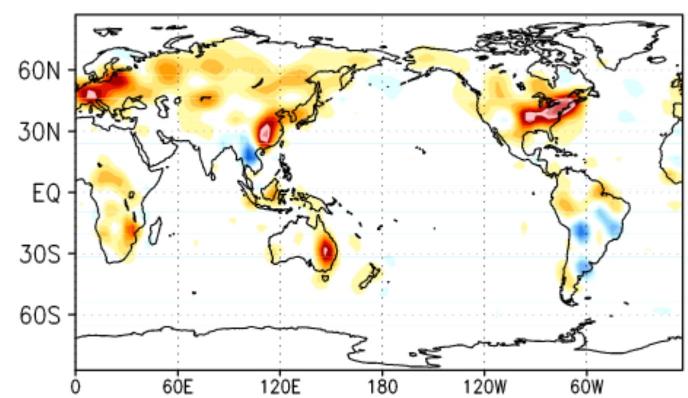



*We succeed in estimating time-evolving CF at model-grid scale*

00Z01JAN ▶
After one year of DA



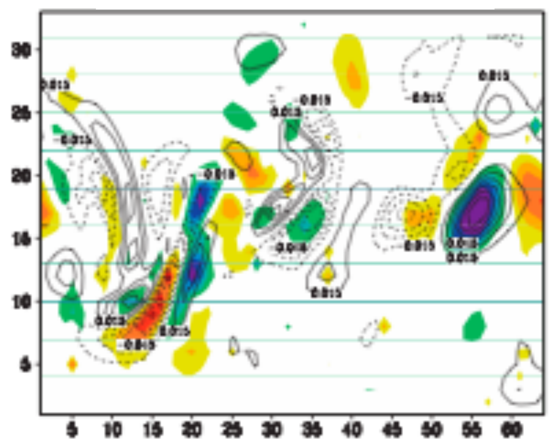
# Summary

- We have shown the feasibility of simultaneous analysis of meteorological and carbon variables within LETKF framework through the 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)
- The implementation of the LETKF-C to NCAR CAM 3.5 model is now in progress
  - Analysis step shows very good performance in OSSE with real observation coverage
  - Analysis cycle with a forecast step will be operated soon
- The same methodology has been applied to estimating surface fluxes of heat, moisture and momentum, and the results are promising!

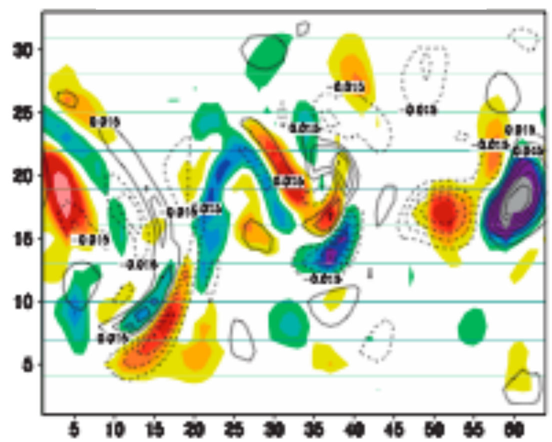
# Initial and final analysis corrections (colors), with one BV (contours)

LETKF

Initial increments



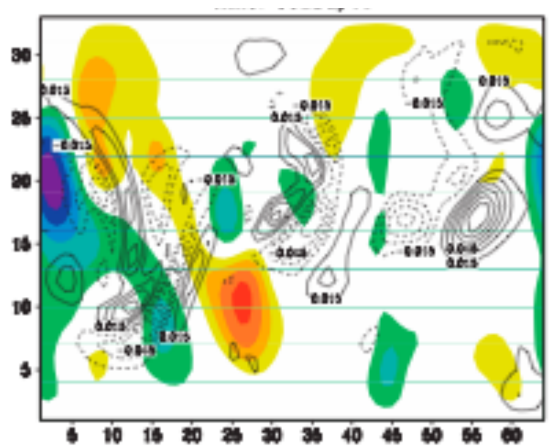
Final increments



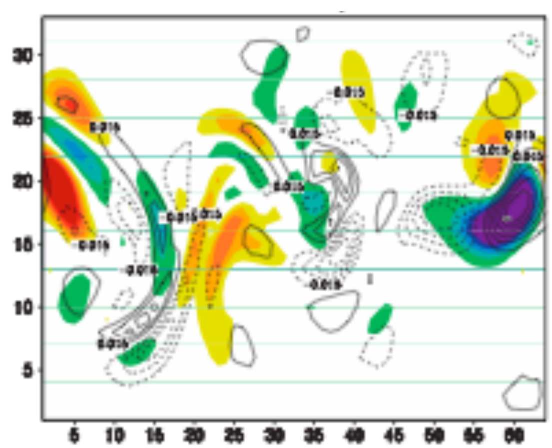
LETKF

4D-Var-12hr

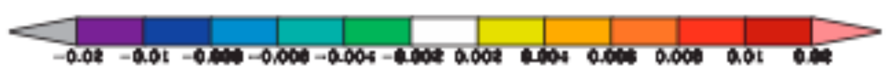
Initial increments



Final increments



4D-Var-12hr



GODAE Ocean View/WGNE Workshop 2013
19 March 2013

# Data assimilation for the coupled ocean-atmosphere

Eugenia Kalnay, Tamara Singleton, Steve Penny, Takemasa Miyoshi, Jim Carton

Thanks to the UMD Weather-Chaos Group, to Daryl Kleist and to the India Monsoon Mission

# Outline

- Traditional approaches.
- Thesis of Tamara Singleton (DA with toy coupled model).
- The LETKF and Running in Place.
- Steve Penny: 7 years ocean reanalysis.
- Steve Penny: New EnKF-based hybrid.
- Shaoqing Zhang: GFDL coupled EnKF.
- Our planned approach to coupled LETKF (India Monsoon Mission)
- Questions:
  - Can we do a robust coupled SST analysis? SSH? Scatterometer winds?
  - Should we do LETKF-RIP? Short windows for the ocean and atm.?
  - Should we do Gaussian Transformation (Lien et al.)
  - Should we do Proactive QC with Ens. Fcst. Sens. to Obs. (EFSO)?
- Discussion

# Traditional approaches

“In a typical coupling scheme for an ocean-atmosphere model, the ocean model passes SST to the atmosphere, while the atmosphere passes back heat flux components, freshwater flux, and horizontal momentum fluxes.” (Neelin, Latif & Jin, 1994)

SST in the ocean model is frequently nudged from Reynolds SSTs, not assimilated from observations. SSH may not be even be used.

The data assimilation windows are very different for the ocean and the atmosphere.

# Tamara Singleton's thesis

Data Assimilation Experiments with a
Simple Coupled Ocean-Atmosphere Model

Questions she addressed:

-- Which is more accurate: 4D-Var or EnKF?

-- Is it better to do an ocean reanalysis separately, or as a single coupled system?

-- ECCO is a version of 4D-Var where both the initial state and the surface fluxes are control variables. This allows ECCO to have very long windows (decades) and estimate the surface fluxes that give the best analysis.

Is ECCO the best approach for ocean reanalysis?

# Simple Coupled Ocean-Atmosphere System

**3 coupled Lorenz models**: A slow "ocean" component strongly coupled with a fast "tropical atmosphere component", in turn weakly coupled with a fast "extratropical atmosphere" (Peña and Kalnay, 2004).

**Model Parameter Definitions**

| Variables | Description | Values |
|---|---|---|
| $c, c_z, c_e$ | Coupling coefficient | $c, c_z = 1$<br>$c_e = 0.08$ |
| $\tau$ | time scale | $\tau = 0.1$ |
| $\sigma$, $b$, and $r$ | Lorenz parameters | $\sigma=10$, $b=8/3$, and $r=28$ |
| $k_1, k_2$ | Uncentering parameters | $k_1=10$<br>$k_2=-11$ |

**Extratropical atmosphere**

$$\dot{x}_e = \sigma(y_e - x_e) - c_e(x_t + k_1)$$
$$\dot{y}_e = rx_e - y_e - x_e z_e - c_e(y_t + k_1)$$
$$\dot{z}_e = x_e y_e - b z_e$$

**Tropical atmosphere**

$$\dot{x}_t = \sigma(y_t - x_t) - c(X + k_2) - c_e(x_e + k_1)$$
$$\dot{y}_t = rx_t - y_t - x_t z_t + c(Y + k_2) + c_e(y_e + k_1)$$
$$\dot{z}_t = x_t y_t - b z_e + c_z Z$$

**Ocean**

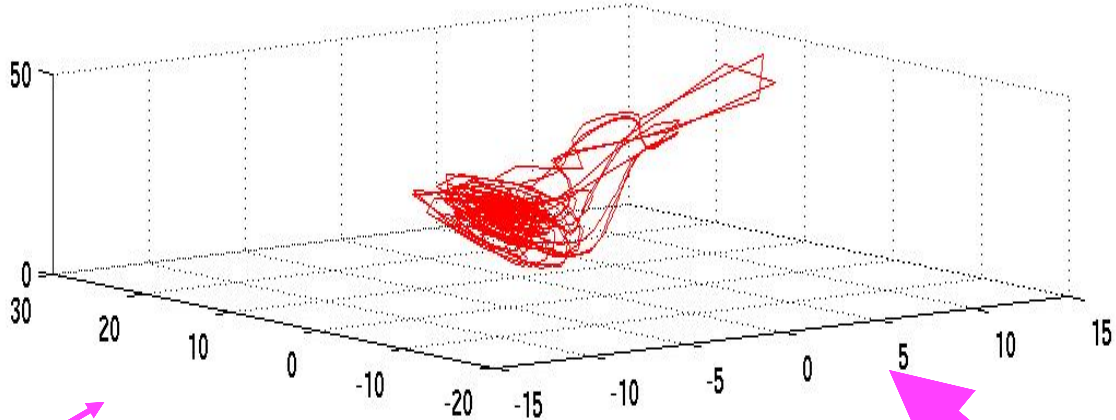
$$\dot{X} = \tau\sigma(Y - X) - c(x_t + k_2)$$
$$\dot{Y} = \tau r X - \tau Y - \tau XZ + c(y_t + k_2)$$
$$\dot{Z} = \tau XY - \tau b Z + c_z z_t$$

Model State: $[x_e, y_e, z_e, x_t, y_t, z_t, X, Y, Z]^T$

# Simple Coupled Ocean-Atmosphere Model (Peña and Kalnay, 2004)

**Tropical Atmosphere**

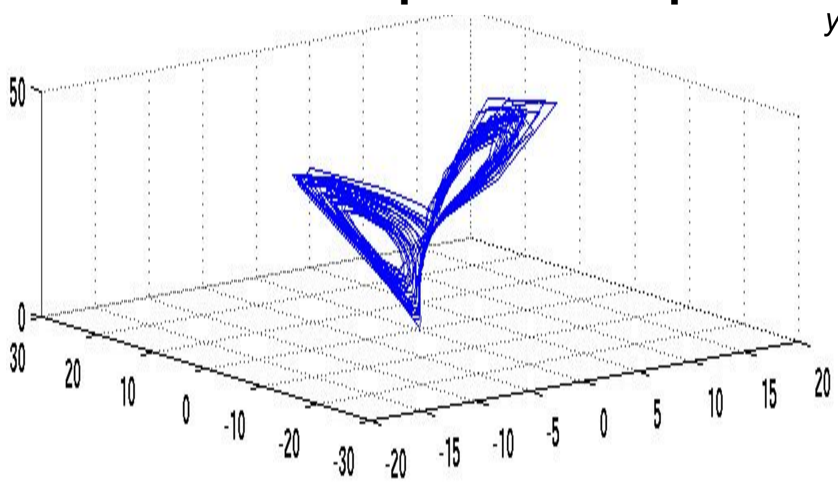
Coupling strength

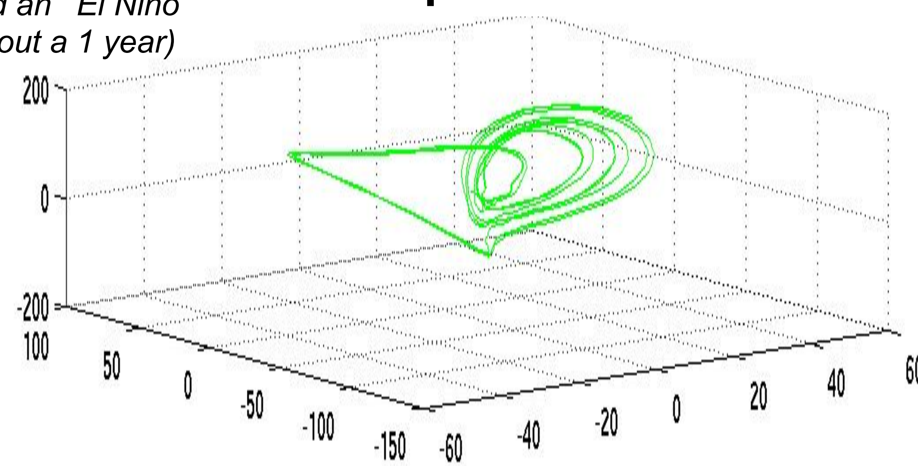



**Extra-tropical Atmosphere**

*Ocean is vacillating between a "normal year" (lasts from ~3-8 years) and an "El Nino" (lasts about a 1 year)*

**Tropical Ocean**

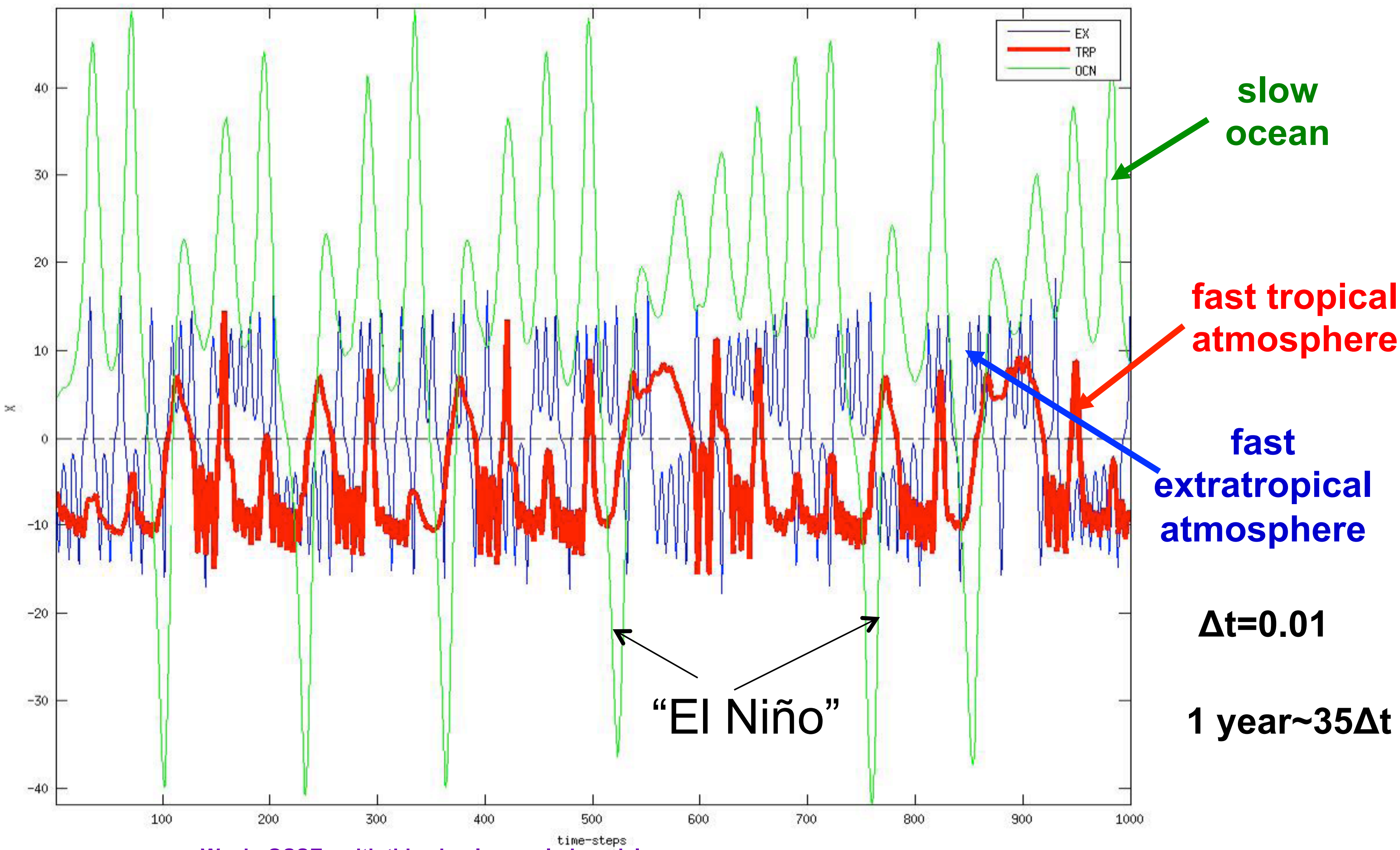



**We do OSSEs with this simple coupled model**

# Simple Coupled Ocean-Atmosphere Model (Peña and Kalnay, 2004)

**Time series of the x-component**




**Δt=0.01**

**1 year~35Δt**

We do OSSEs with this simple coupled model

# 4D-Var/ETKF Data Assimilation Summary

- **We developed a 4D-Var data assimilation system for the simple coupled ocean-atmosphere model**
- We found that **lengthening the assimilation window** and **applying QVA improves the 4D-Var analysis**.
- **Tuning the amplitude of the background error covariance has an impact on the performance of the assimilation**.
- **EnKF-based methods (LETKF & ETKF-QOL) compete with 4D-Var analyses** for short and long assimilation windows.
- For much **longer assimilation windows, 4D-Var outperforms the EnKF-based methods**

- **Short windows are good for ETKF**
- **Long windows are good for 4D-Var**
- **Optimal accuracy similar for 4D-Var and ETKF**

# ECCO-like 4D-Var

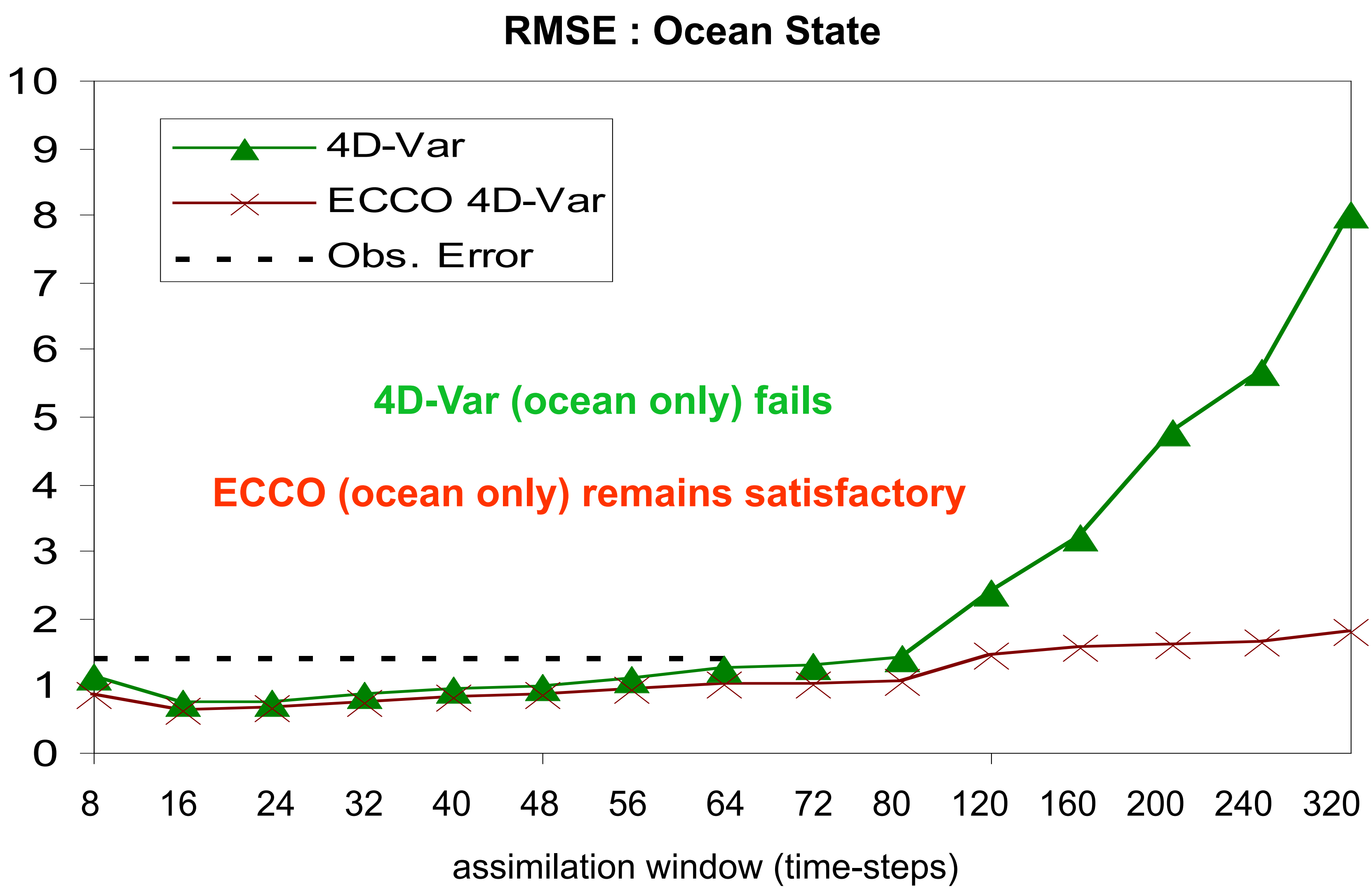
- The consortium for Estimating the Circulation and Climate of the Ocean (ECCO) is a collaboration of a group of scientists from the MIT, JPL, and the Scripps Institute of Oceanography
- The main characteristic of ECCO is that they include surface fluxes as control variables.
  - This allows them to have exceedingly long assimilation windows in 4D-Var (e.g. 10 years or even 50 years).
  - They used NCEP Reanalysis fluxes (Kalnay et al, 1996) as a first guess.
- ECCO used 4D-Var to estimate the initial ocean state and surface fluxes (Stammer et al., 2004; Kohl et al., 2007) in a 50-year reanalysis

# Comparison of ECCO-like & Ocean 4D-Var

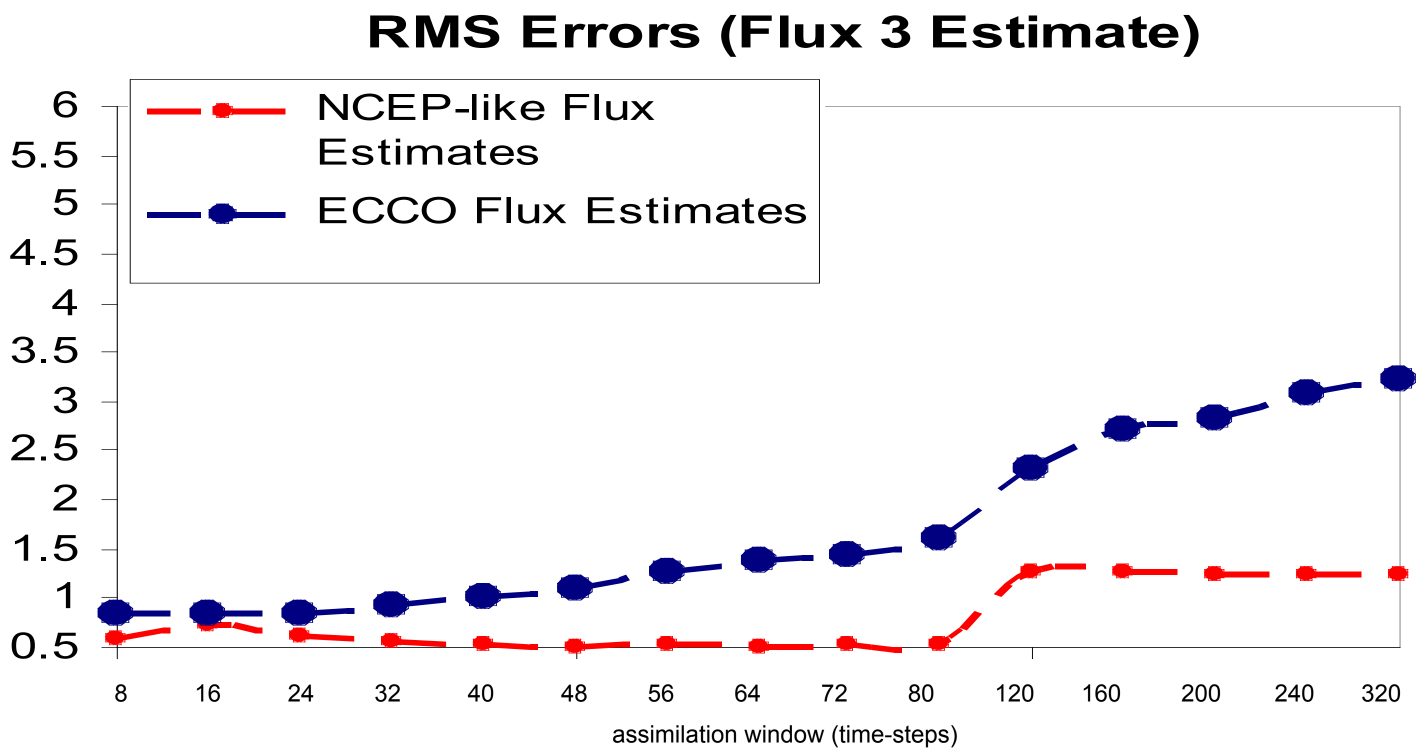
QVA APPLIED

OCEAN ONLY

Obs. s.d error = 1.41 for ocean



# Are the ECCO fluxes more accurate?




**ECCO does not improve the flux estimates**

# Answers to the Research Questions

**Questions:**

-- Which is more accurate: 4D-Var or EnKF?

Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

# Answers to the Research Questions

**Questions:**

-- Which is more accurate: 4D-Var or EnKF?

Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

-- Is it better to do the ocean reanalysis separately, or as a single coupled system?

Both EnKF and 4D-Var are similar and most accurate when coupled, but uncoupled (ocean only) reanalyses are fairly good.

# Answers to the Research Questions

**Questions:**

-- Which is more accurate: 4D-Var or EnKF?
Fully coupled EnKF (with short windows) and 4D-Var (with long windows) have about the same accuracy.

-- Is it better to do the ocean reanalysis separately, or as a single coupled system?
Both EnKF and 4D-Var are similar and most accurate when coupled, but uncoupled (ocean only) reanalyses are quite good.

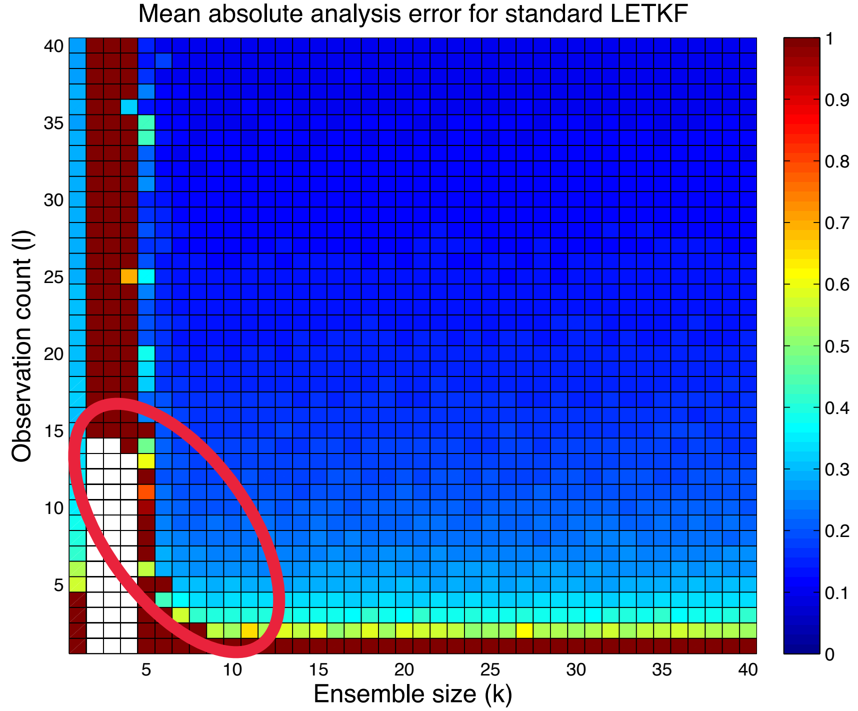
-- Is ECCO 4D-Var with both the initial state and the surface fluxes as control variables the best approach?
In our simple ocean model 4D-Var cannot remain accurate with very long windows. Our ECCO reanalysis remained satisfactory with very long windows but at the expense of less accurate fluxes.

# How about hybrids between Var and EnKF?

- So far hybrids have been created combining an existing Var system with an ensemble to provide the flow dependence of the background error covariance.
- **We would like to start with a well-developed EnKF (like the LETKF) and add a 3D-Var that provides the full rank that the ensemble lacks.**
- Steve Penny developed a simple, locally Gaussian 3D-Var for this purpose, and tested it on the Lorenz-96, a 40 variable model.
- He plots the analysis error as a function of the number of ensemble members (2 to 40) and the number of observations (1 to 40).

# An ensemble based hybrid with a simple local 3D-Var (Steve Penny) applied to the Lorenz 96 model

Standard LETKF

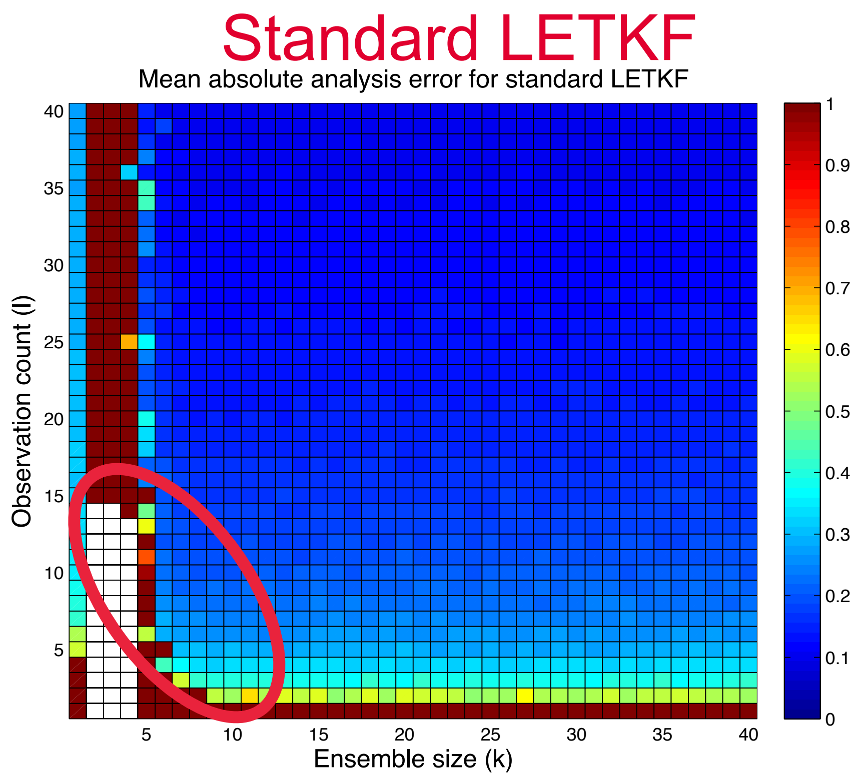


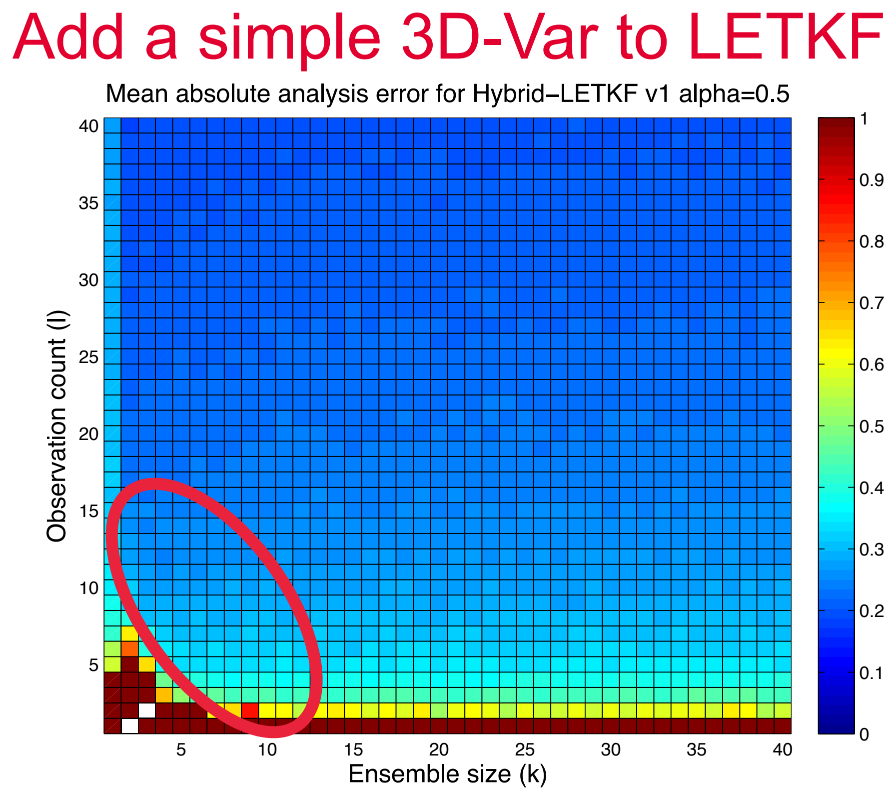

This is the corner where we are in ocean EnKF: too few obs, too few ensembles

The total model dimension is K=40

The LETKF is extremely accurate as long as k>7, number of obs>7.

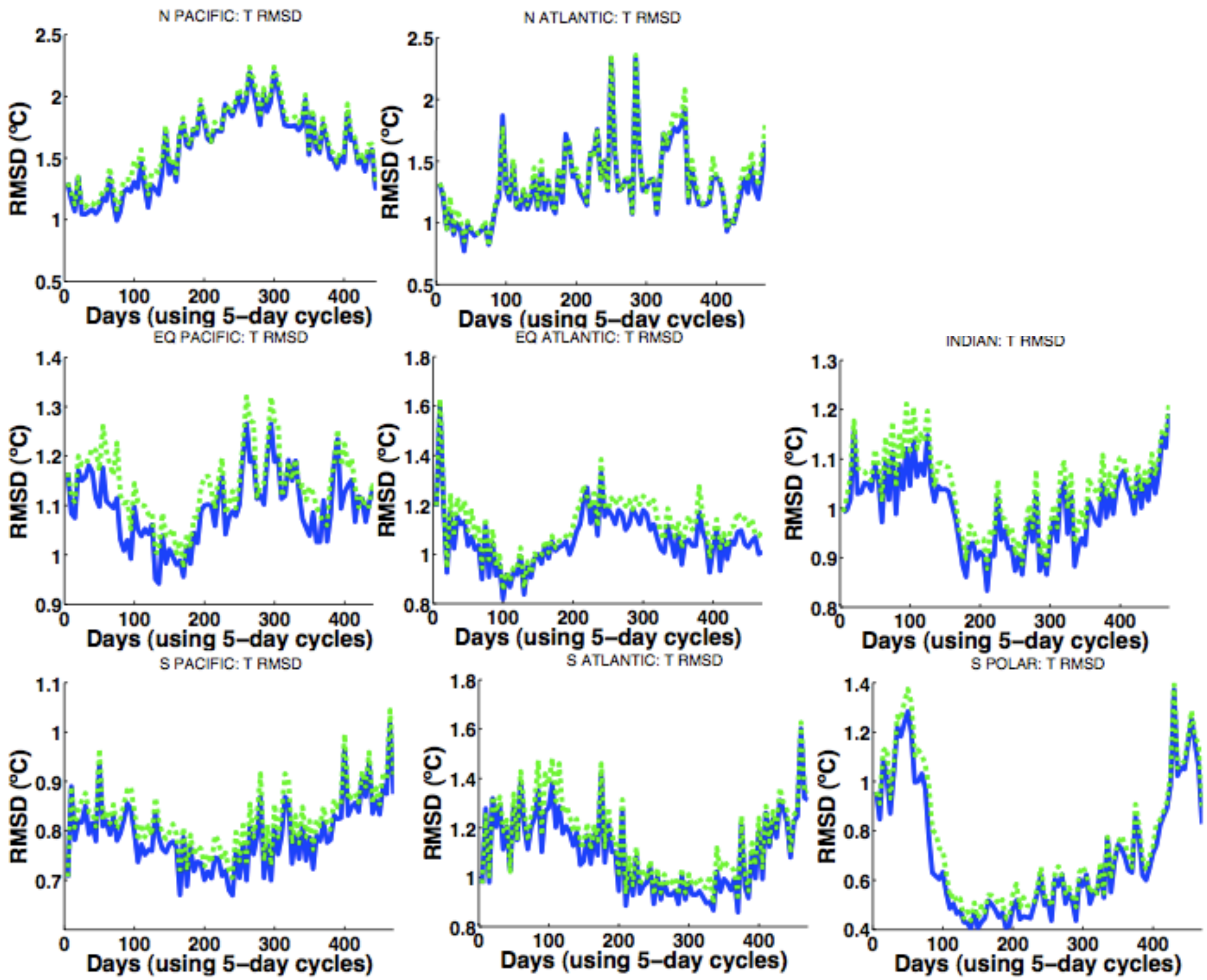
# An ensemble based hybrid with a simple local 3D-Var (Steve Penny) applied to the Lorenz 96 model



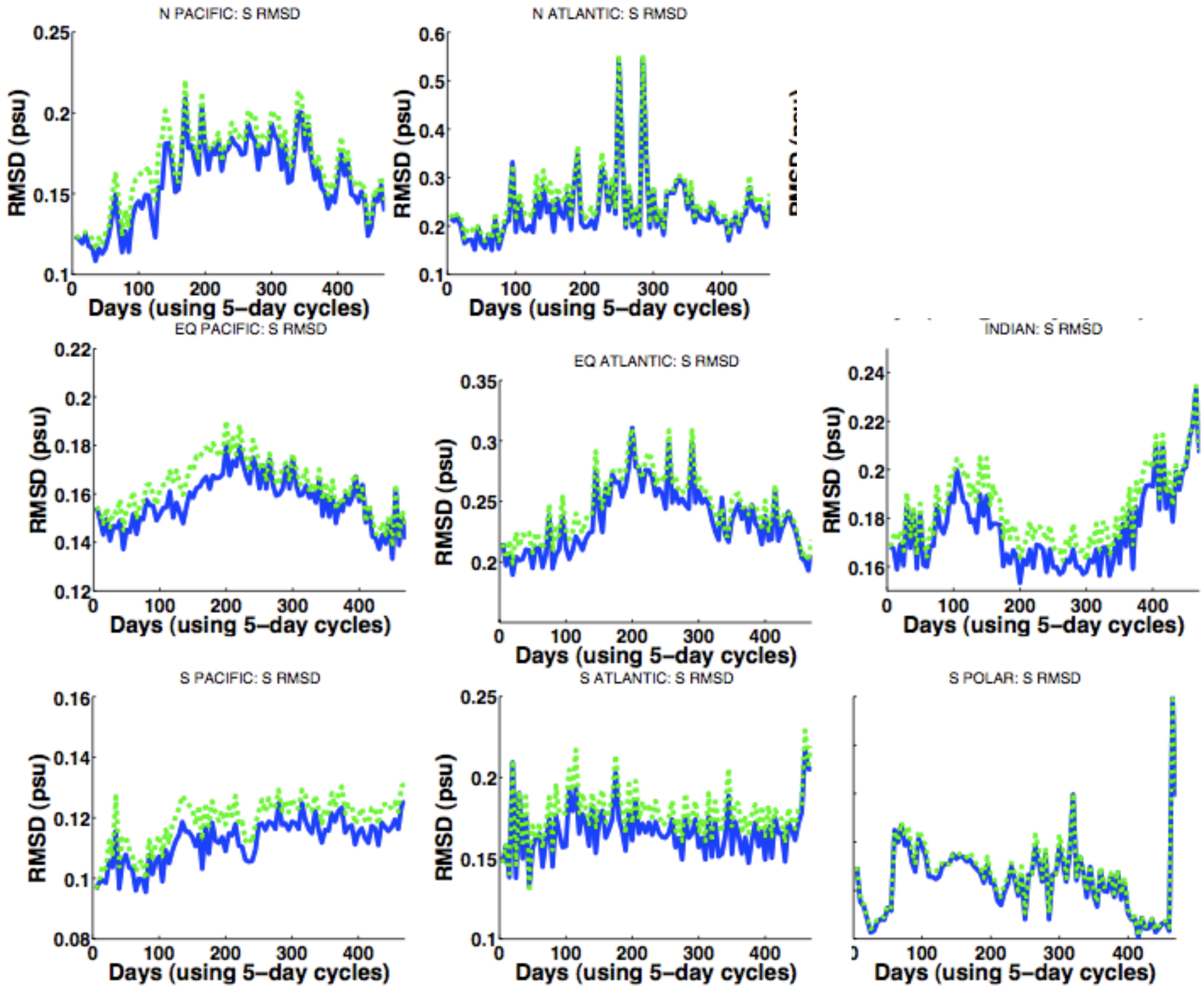

**The hybrid LETKF-simple 3D-Var is more robust for few ensemble members and few observations, as in the ocean.**

# Penny's new ocean hybrid reanalysis: LETKF + GODAS hybrid



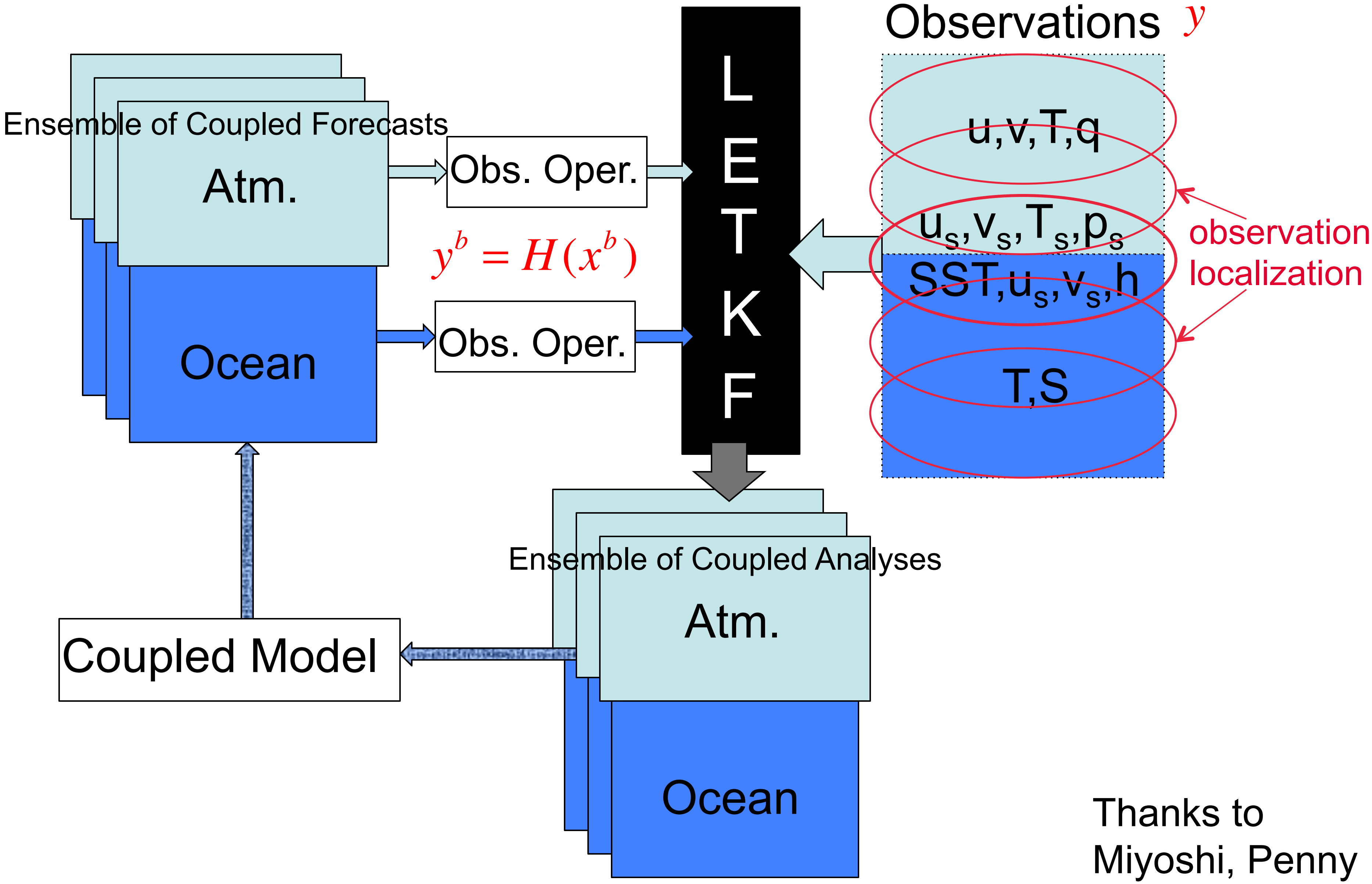

Temperature O-F RMSD

# Penny's new ocean hybrid reanalysis: LETKF + GODAS hybrid




Salinity O-F RMSD

# Basic idea for our coupled LETKF assimilation




Thanks to
Miyoshi, Penny

# Summary: ideas/questions for future coupled ocean-atmosphere EnKF

- Toy model: coupled assimilation and short windows are more accurate for LETKF even if ocean has longer time scales.
- Running in Place (RIP) extracts more information from the observations and allows the use of shorter windows.
- A new hybrid LETKF+simple 3D-Var would make the system more robust with fewer ensemble members and observations.
- For the coupled (India Monsoon Mission) CFS system, we will test the use of 6hr (short) windows for the ocean as well as the atmosphere assimilation.
- Assimilate SST and SSH observations directly.
- Localization of observations near the surface should allow for atm.-ocean interaction through the background error covariance

# Summary: ideas/questions for future coupled ocean-atmosphere EnKF

- Toy model: coupled assimilation and short windows are more accurate for LETKF even if ocean has longer time scales.
- Running in Place (RIP) extracts more information from the observations and allows the use of shorter windows.
- A new hybrid LETKF+simple 3D-Var would make the system more robust with fewer ensemble members and observations.
- For the coupled (India Monsoon Mission) CFS system, we will test the use of 6hr (short) windows for the ocean as well as the atmosphere assimilation.
- Assimilate SST and SSH observations directly.
- Localization of observations near the surface should allow for atm.-ocean interaction through the background error covariance

**Thanks!**