Recent Advances in EnKF

Former students (Shu-Chih Yang, Takemasa Miyoshi, Hong Li, Junjie Liu, Chris Danforth, Ji-Sun Kang, Matt Hoffman, Steve Penny, Steve Greybush, Tamara Singleton, Javier Amezcua), and

Eugenia Kalnay

University of Maryland

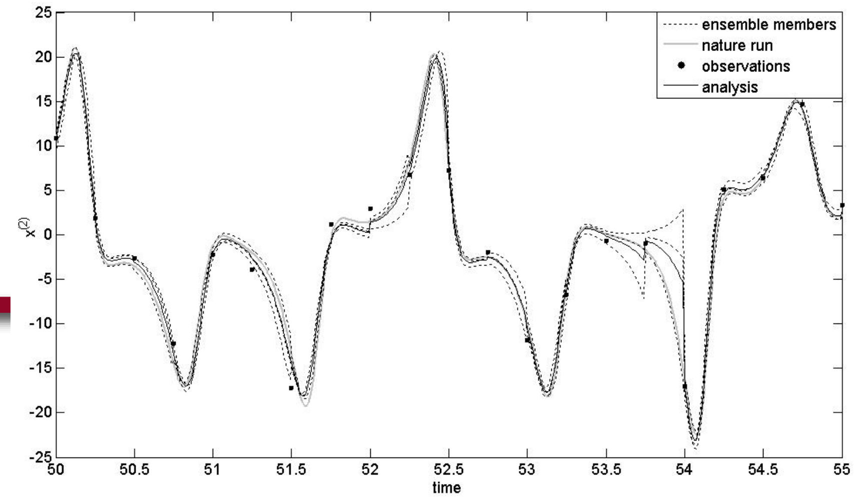
UMD Weather-Chaos Group: **Kayo Ide**, **Brian Hunt**, Ed Ott,

and students (Guo-Yuan Lien, Yan Zhou, Adrienne Norwood, Erin Lynch, Yongjing Zhao, Daisuke Hotta)

Also: Yoichiro Ota, Malaquías Peña, Matteo Corazza.

JCSDA, 27 July 2012

Data Assimilation



- Combine "optimally" short-term forecasts with observations.
- 3D-Var and Optimal Interpolation: used for many years, fixed background error covariance **B**
- Advanced methods: they evolve **B** ("errors of the day"):
  - ✓ 4D-Var: widely used in operations.

    Full rank **B** :-); requires model adjoint. :-(
  - ✓ Ensemble Kalman Filter, low rank **B** :-(; no adjoint :-)
  - ✓ Hybrids: **B** from EnKF, variational solution

Conclusions from the THORPEX Workshop in Buenos Aires (2008)

- ✓ 4D-Var and EnKF are competitive in skill
- ✓ Hybrid approach best (Buehner et al, 2008, 2009)
- ✓ There are no fatal disadvantages for either system
- ✓ Computationally competitive
- ✓ About 40-100 ensemble members needed from storm to global scales for EnKF
- ✓ Both methods have developed approaches to deal with model errors and nonlinearities

As a result, UKMO, NCEP, ECMWF, Italy, Japan, Canada, Germany, Brazil, Argentina… are now exploring EnKF (or hybrid EnKF+variational) for operations. NCEP implemented a hybrid on May 22, 2012!

This talk: tools that improve EnKF

We adapted ideas inspired by 4D-Var:

✓ **No-cost smoother** (Kalnay et al, Tellus 2007)

✓ “**Running in Place**” and “**Quasi Outer Loop**”, deal with **spin-up**, nonlinearities and long windows (K. & Yang, QJ 2010, Yang et al. MWR 2012)

✓ **Forecast sensitivity** to observations (Liu and Kalnay, QJ, 2008). **Ota et al.**

✓ **Coarse** analysis resolution **without degradation** (Yang et al., QJ, 2009)

✓ Low-dimensional **model bias correction** (Li et al., MWR, 2009)

✓ Simultaneous estimation of **optimal inflation** and **observation errors** (Li et al., QJ, 2009).

Examples of applications:

✓Global Ocean Data Assimilation (Penny, PhD thesis, 2011)

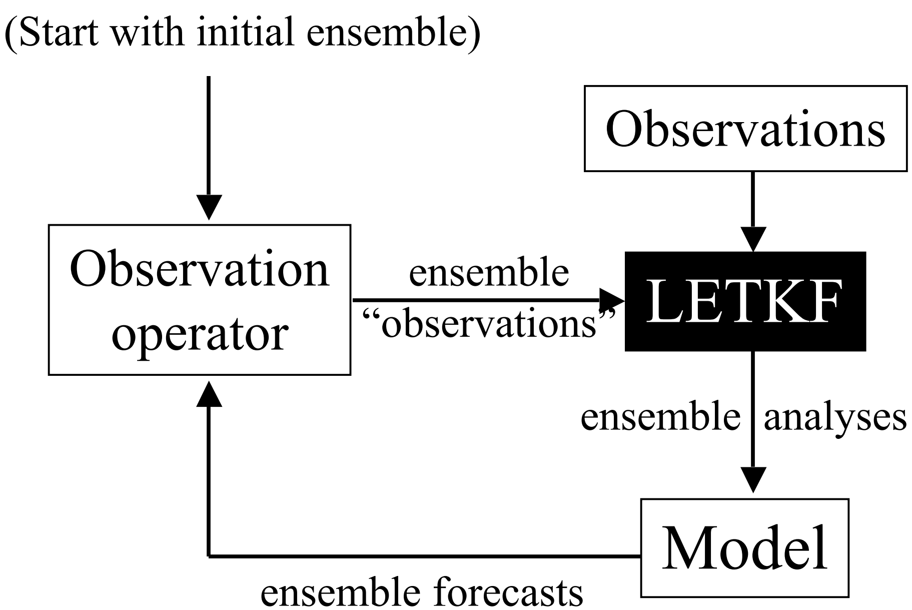
✓Estimates of *surface carbon fluxes as parameters* (Kang et al, 2011)

Comparison of EnKF/4DVar/ECCO in a simple coupled ocean-atm model (Singleton, 2011).

✓ **Initial** and **final** increments in 4D-Var and EnKF

✓ Assimilation of precipitation (Lien et al., 2012)

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

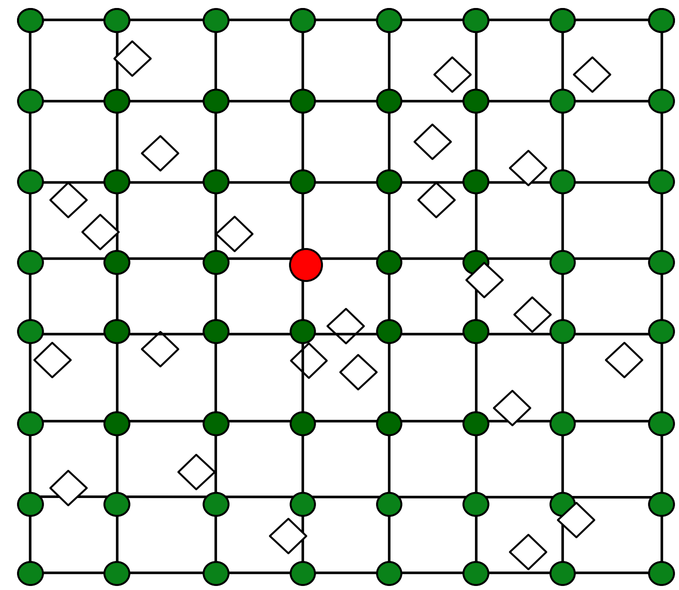


- Model independent (black box)
- Obs. assimilated simultaneously at each grid point
- 100% parallel
- No **adjoint** needed
- 4D LETKF extension
- Computes **weights** explicitly

Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

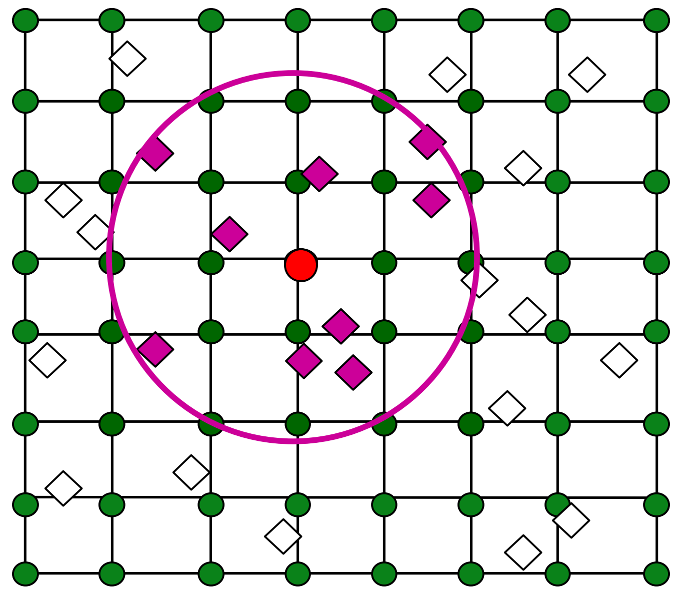


Localization based on observations

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:

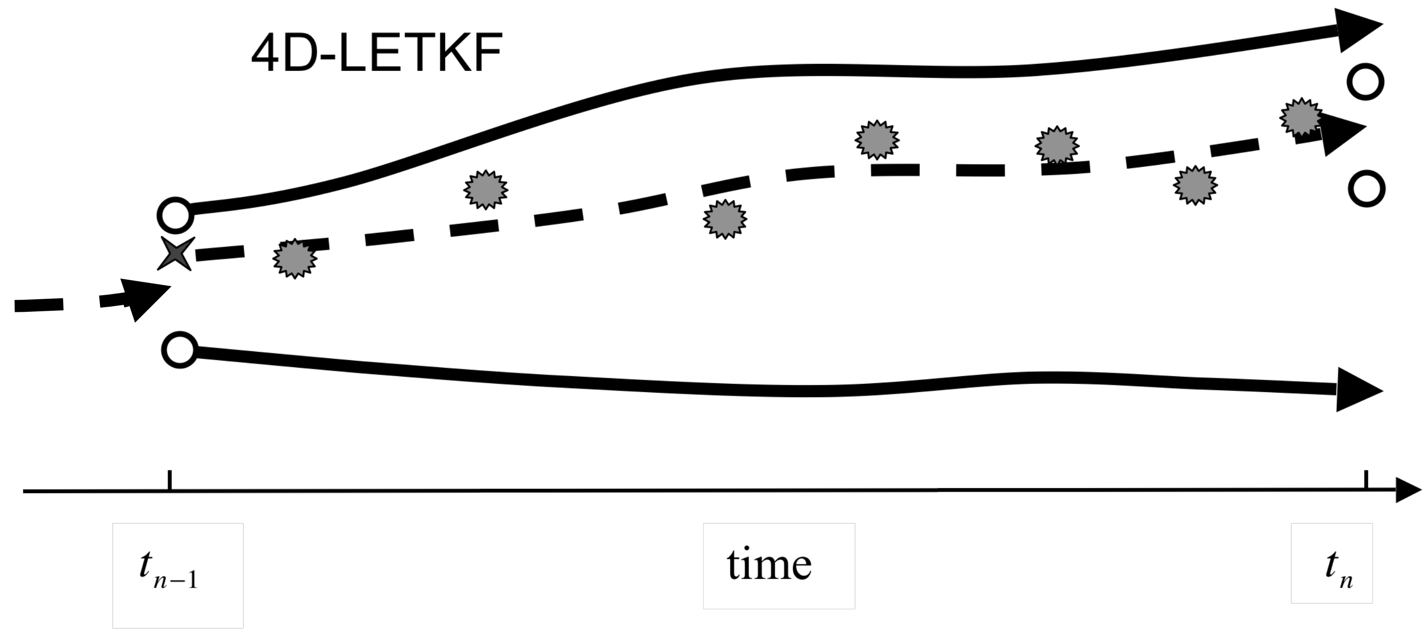
$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1}; \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

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

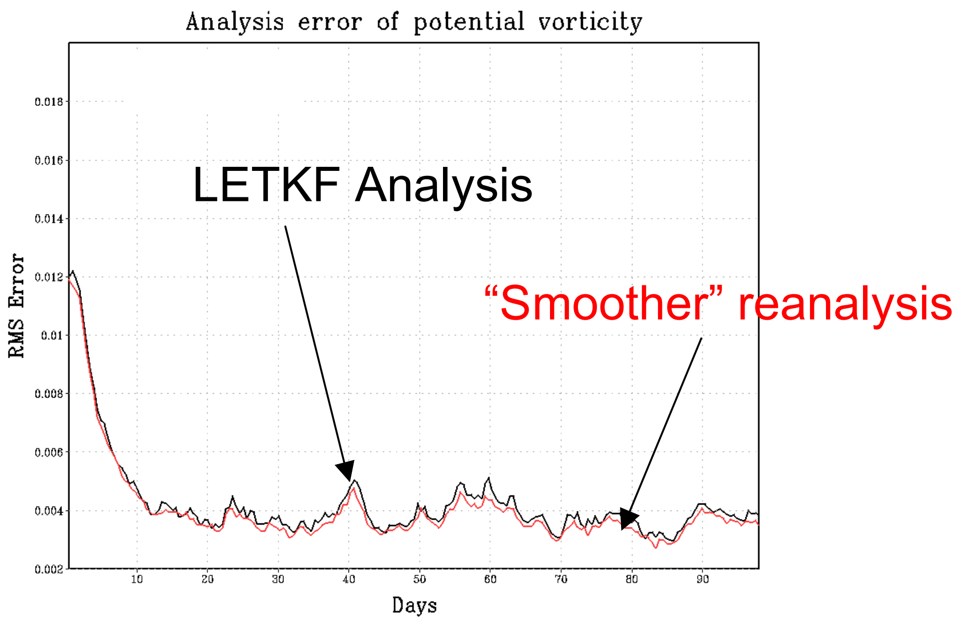
No-cost LETKF smoother tested on a QG model: it works…

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

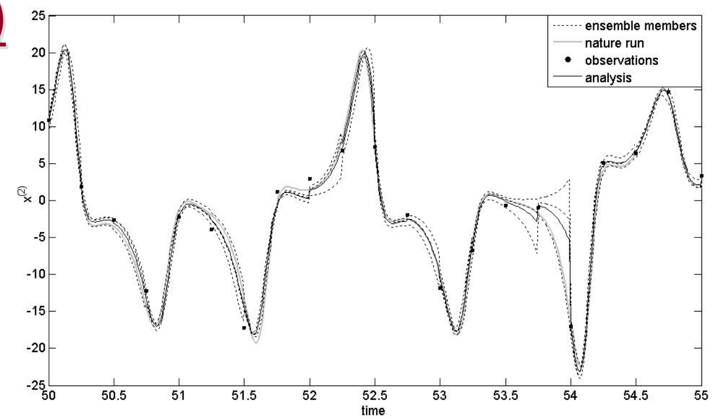
$$\tilde{\mathbf{x}}_{n-1}^a = \overline{\mathbf{x}}_{n-1}^f + \mathbf{X}_{n-1}^f \overline{\mathbf{w}}_n^a$$



This very simple smoother allows us to go back and forth in time within an assimilation window: it allows assimilation of **future** data in reanalysis

Nonlinearities and "outer loop"

- The main disadvantage of EnKF is that it cannot handle nonlinear (non-Gaussian) perturbations and therefore needs short assimilation windows.
- It doesn't have the **outer loop** so important in 3D-Var and 4D-Var (DaSilva, pers. comm. 2006)

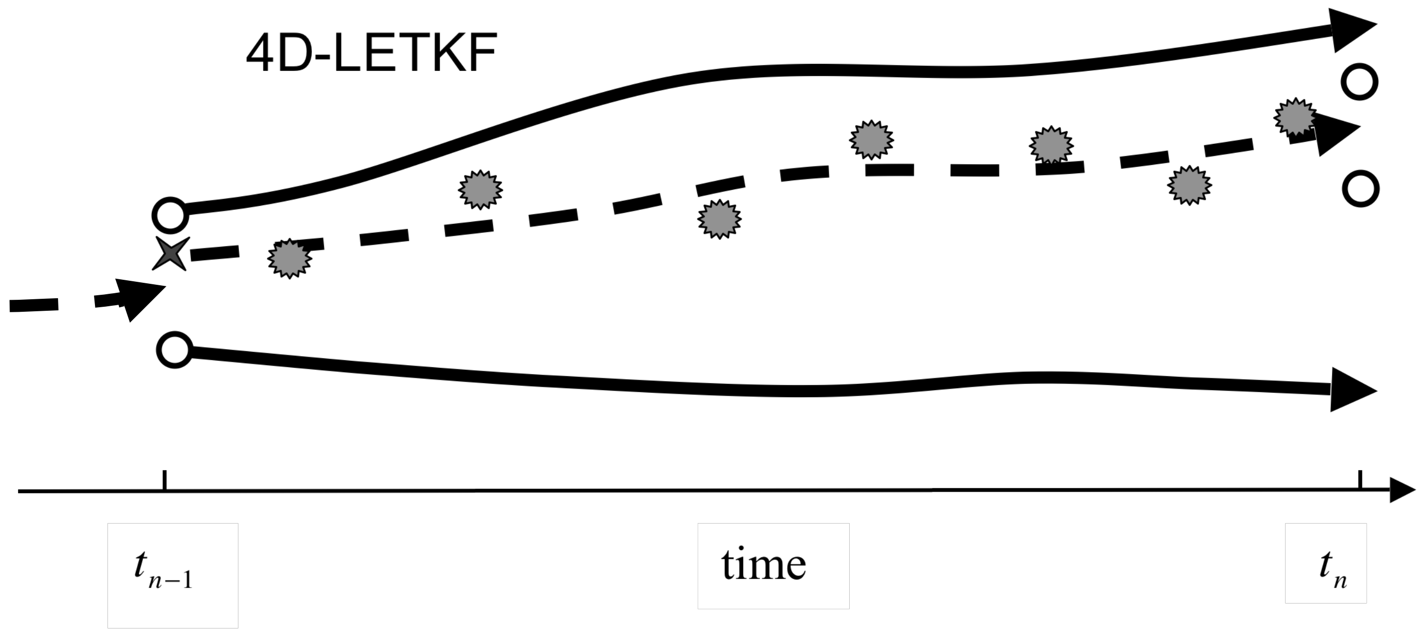


Lorenz -3 variable model (Kalnay et al. 2007a Tellus), RMS analysis error:

	4D-Var	LETKF
Window=8 steps	0.31	**0.30** (linear window)
Window=25 steps	**0.53**	0.66 (nonlinear window)

With long windows + Pires et al. => 4D-Var clearly wins!

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



Quasi Outer Loop (QOL): correct the analysis mean at t_{n-1}

Running in Place (RIP): correct all the analyses at t_{n-1}

…and then do the data assimilation to t_n again

Nonlinearities: “Quasi Outer Loop” (QOL)

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 re-centers the ensemble on a more accurate nonlinear solution.

Lorenz -3 variable model RMS analysis error

	4D-Var	LETKF	LETKF +QOL
Window=8 steps	0.31	0.30	**0.27**
Window=25 steps	0.53	0.66	**0.48**

Nonlinearities, “QOL” and “Running in Place”

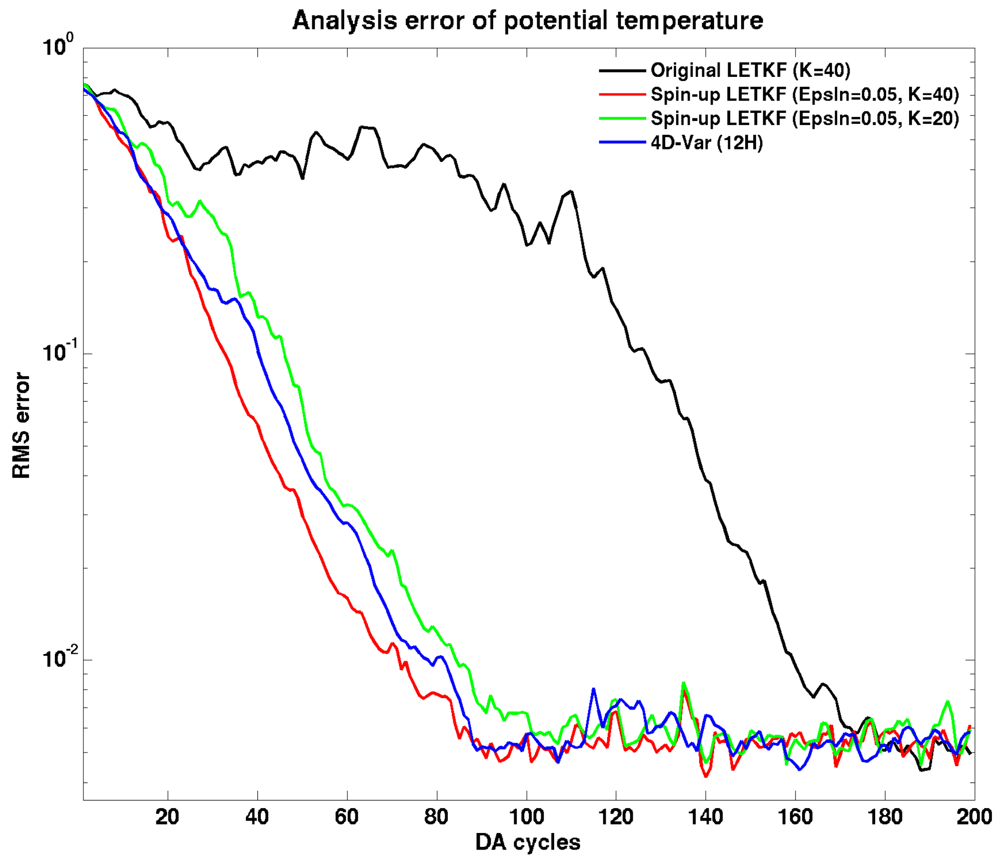
Quasi Outer Loop: similar to 4D-Var: use the final weights to correct only the mean initial analysis, keeping the initial perturbations. Repeat the analysis once or twice. It centers the ensemble on a more accurate nonlinear solution.

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	0.66	**0.48**	**0.39**

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

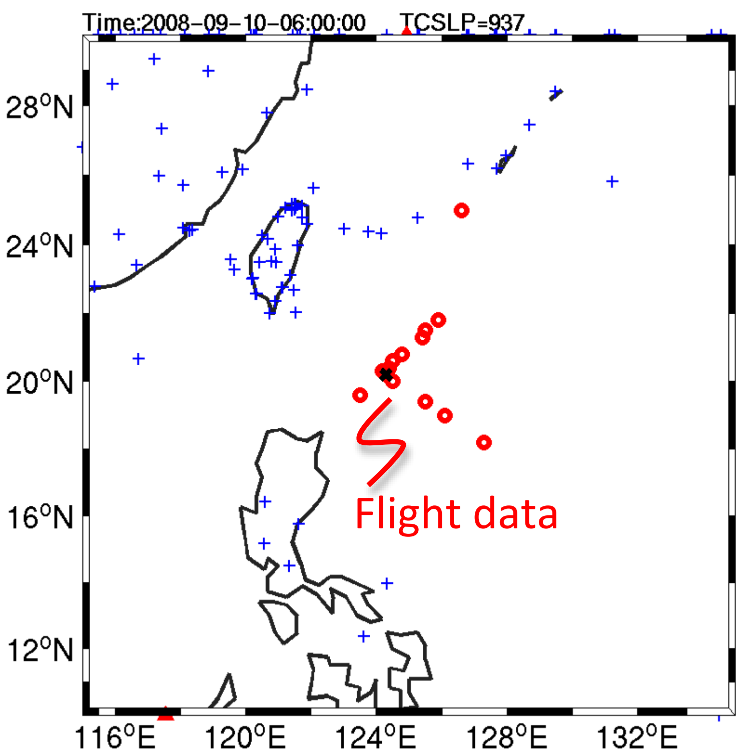
Running in Place: Results with a QG model



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

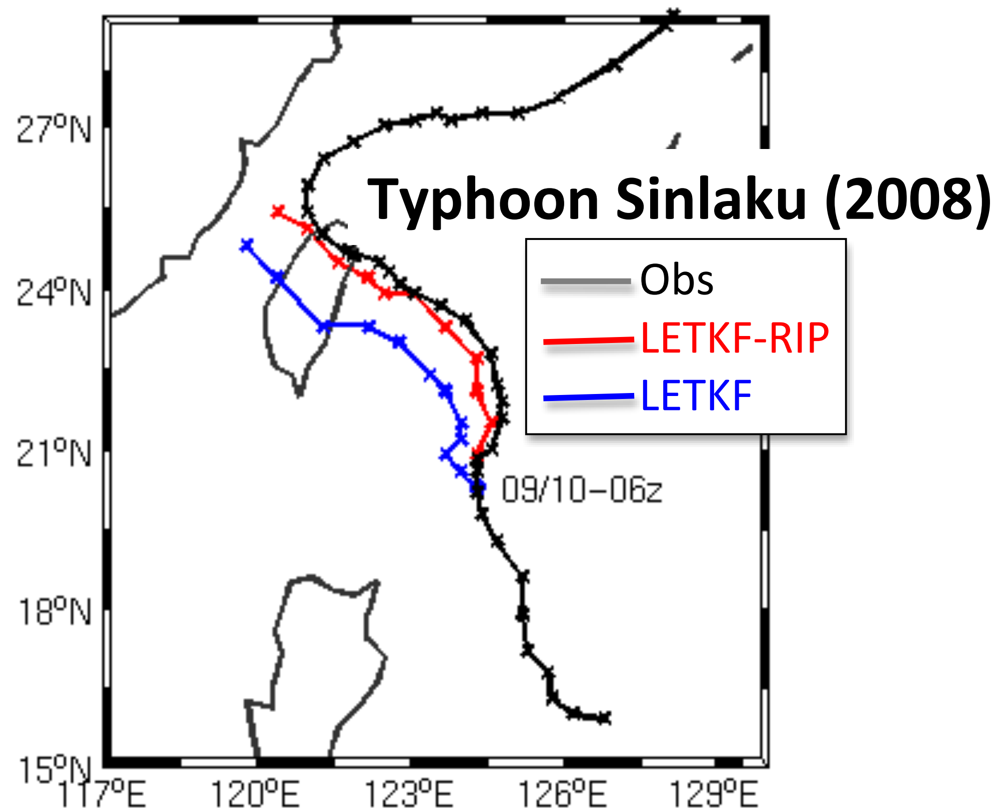
Spin-up depends on initial perturbations, but RIP works well even with random perturbations. It becomes as fast as 4D-Var (blue). RIP takes only 2-6 iterations.

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



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

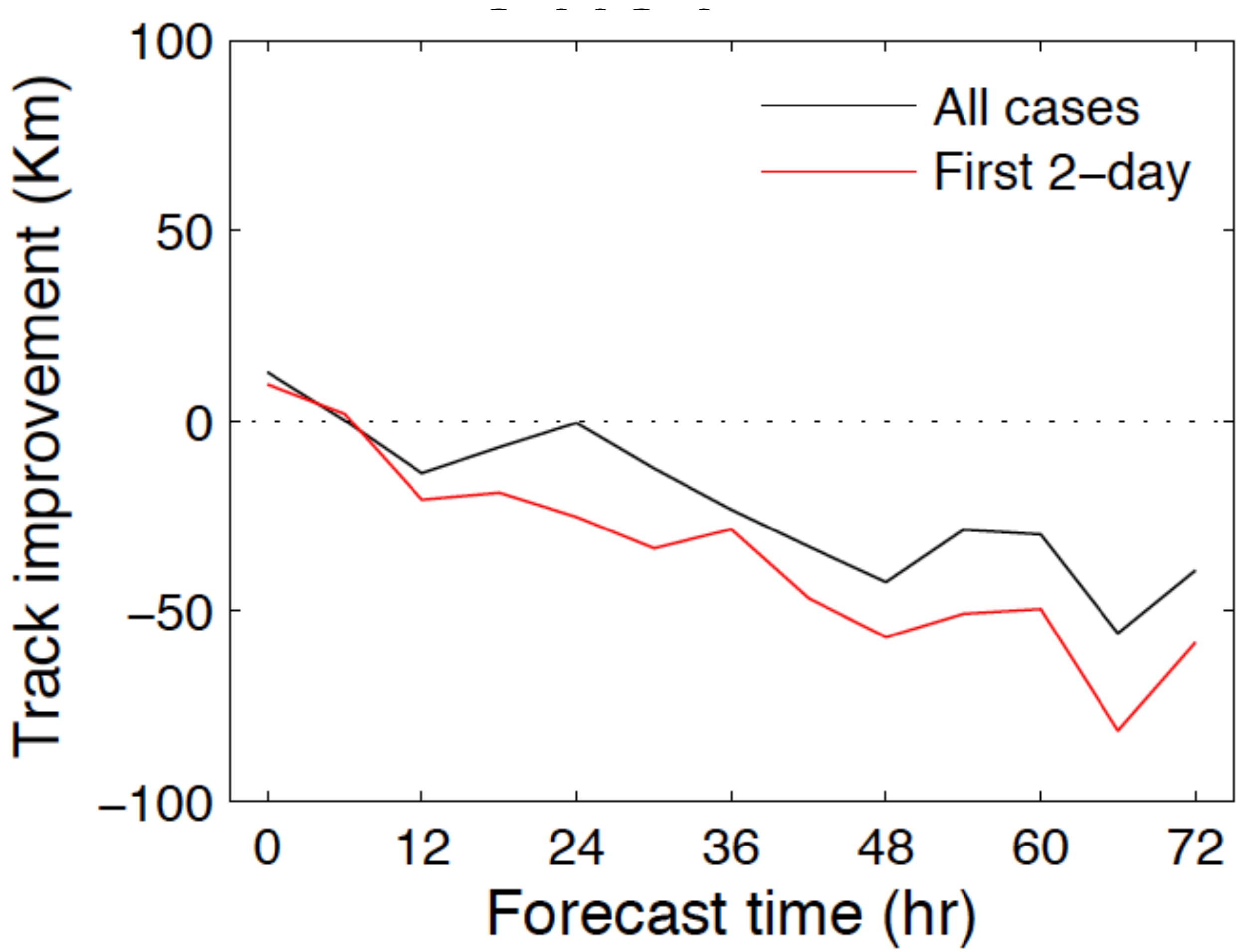
3-day forecast



RIP better use the "limited observations"!

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

Improvement for cross-track



An application of LETKF-RIP to ocean data assimilation

Data Assimilation of the Global Ocean using 4D-LETKF and MOM2

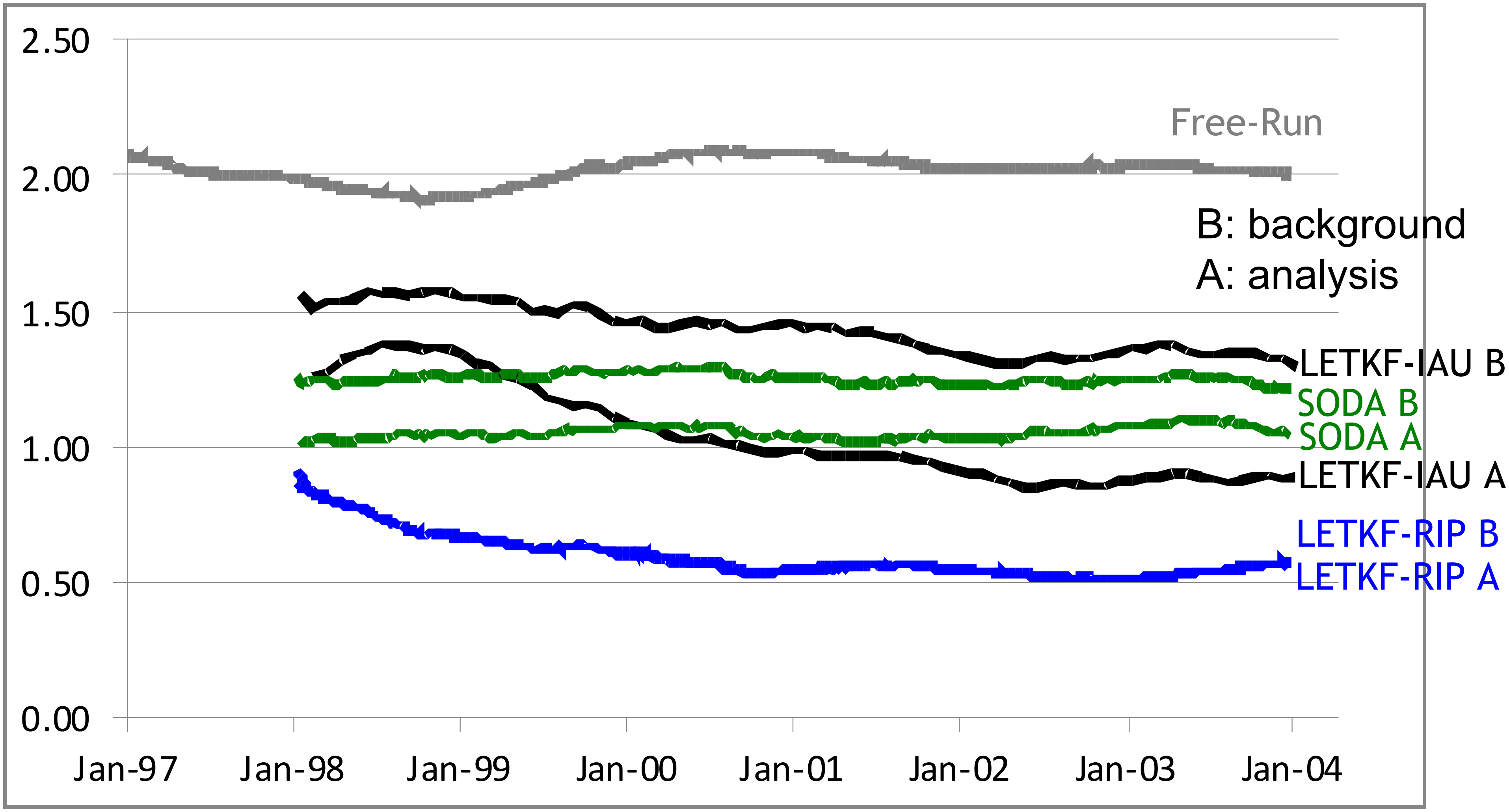
Steve Penny's thesis defense

April 15, 2011

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

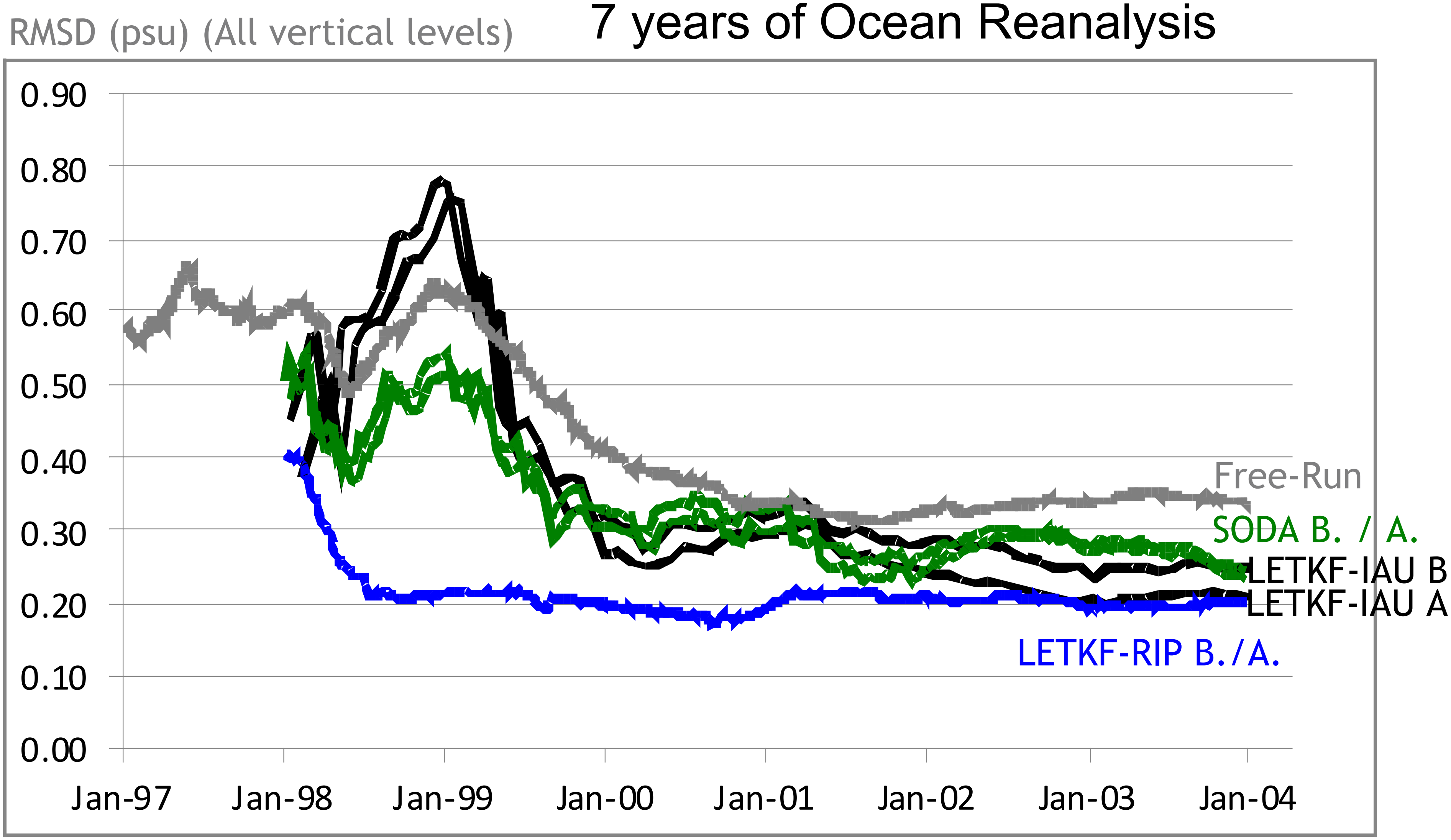
Penny (now at UMD/NCEP) implemented the LETKF with either IAU or RIP and compared it with SODA (OI)

RMSD (°C) (All vertical levels) 7 years of Ocean Reanalysis



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

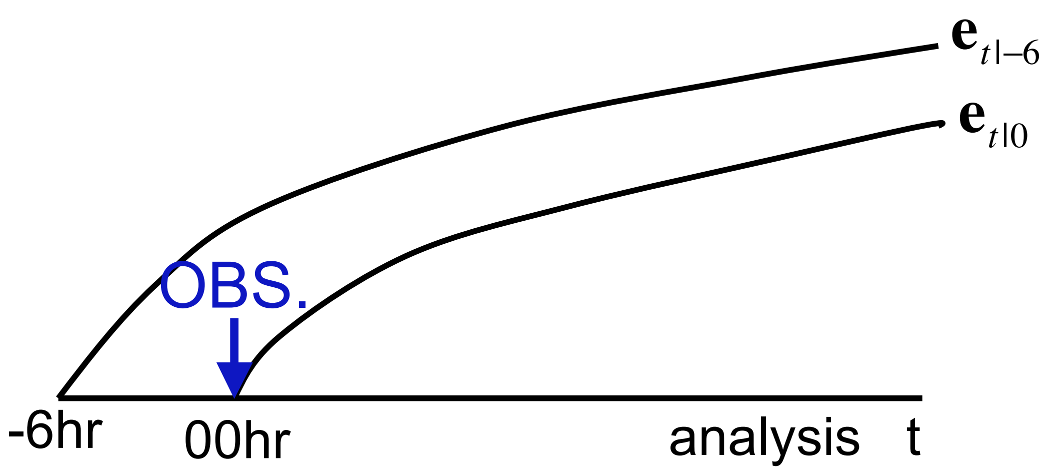
7 years of Ocean Reanalysis



Global RMS(O-F) of Salinity (psu),
12-month moving average
LETKF (with **IAU**), **SODA** and LETKF with **RIP**

Forecast sensitivity to observations

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

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

Forecast sensitivity to observations

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

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

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

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

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

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

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

Forecast sensitivity to observations

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

Langland and Baker (2004):

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

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

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

Liu & Kalnay, Li et al, 2010

This product uses the available **nonlinear** forecast ensemble $\mathbf{X}^{fT}_{t|0}$ and $\mathbf{Y}^a_0 = (\mathbf{HX}^a)$. We can also verify in a targeted area where **P**=1, elsewhere **P**=0:

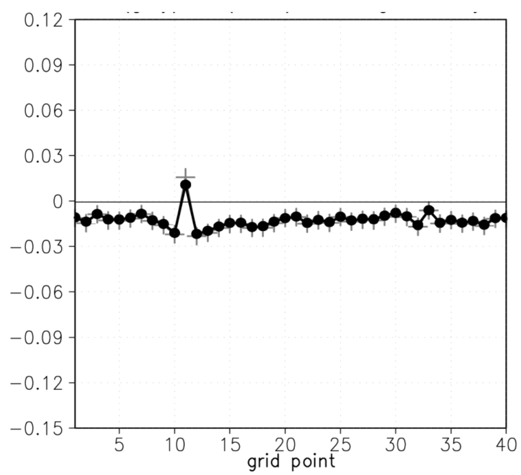
$$\Delta \mathbf{e}^2_P = \left[(\mathbf{y} - H(\mathbf{x}^b_{0|-6}) \right]^T \mathbf{R}^{-1} \mathbf{Y}^a_0 \mathbf{X}^{fT}_{t|0} \mathbf{P}^T \mathbf{P} (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) / (K-1)$$

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

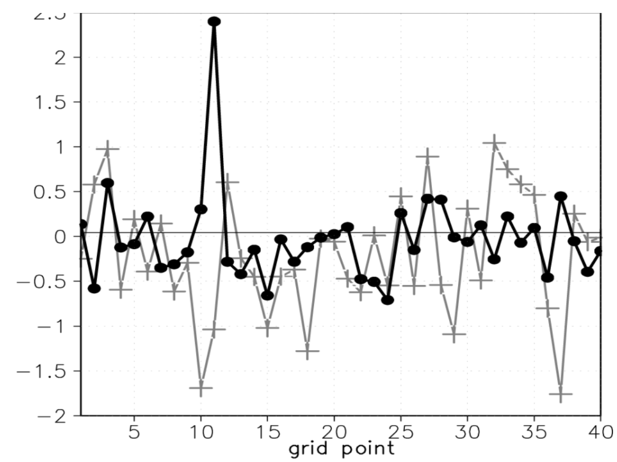
Observation impact from LB(+) and from ensemble sensitivity (●)

(Liu, pers. comm.)

1 day



10 days



✓The adjoint and the ensemble sensitivity give identical observation impact on the 24 hr forecast.

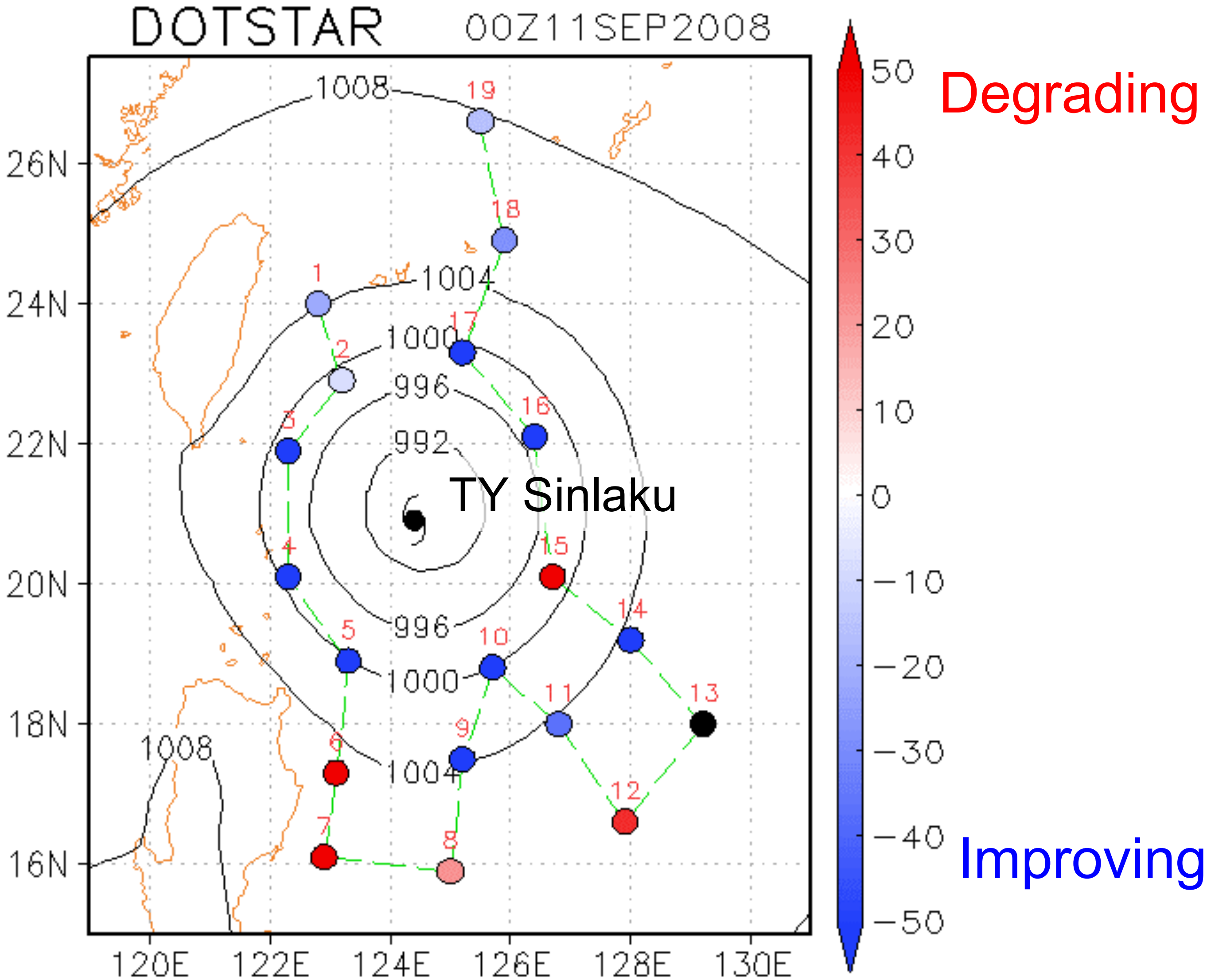
✓The ensemble sensitivity is nonlinear and is able to detect bad obs for longer forecasts.

✓ But we have to deal with localization for longer forecasts.

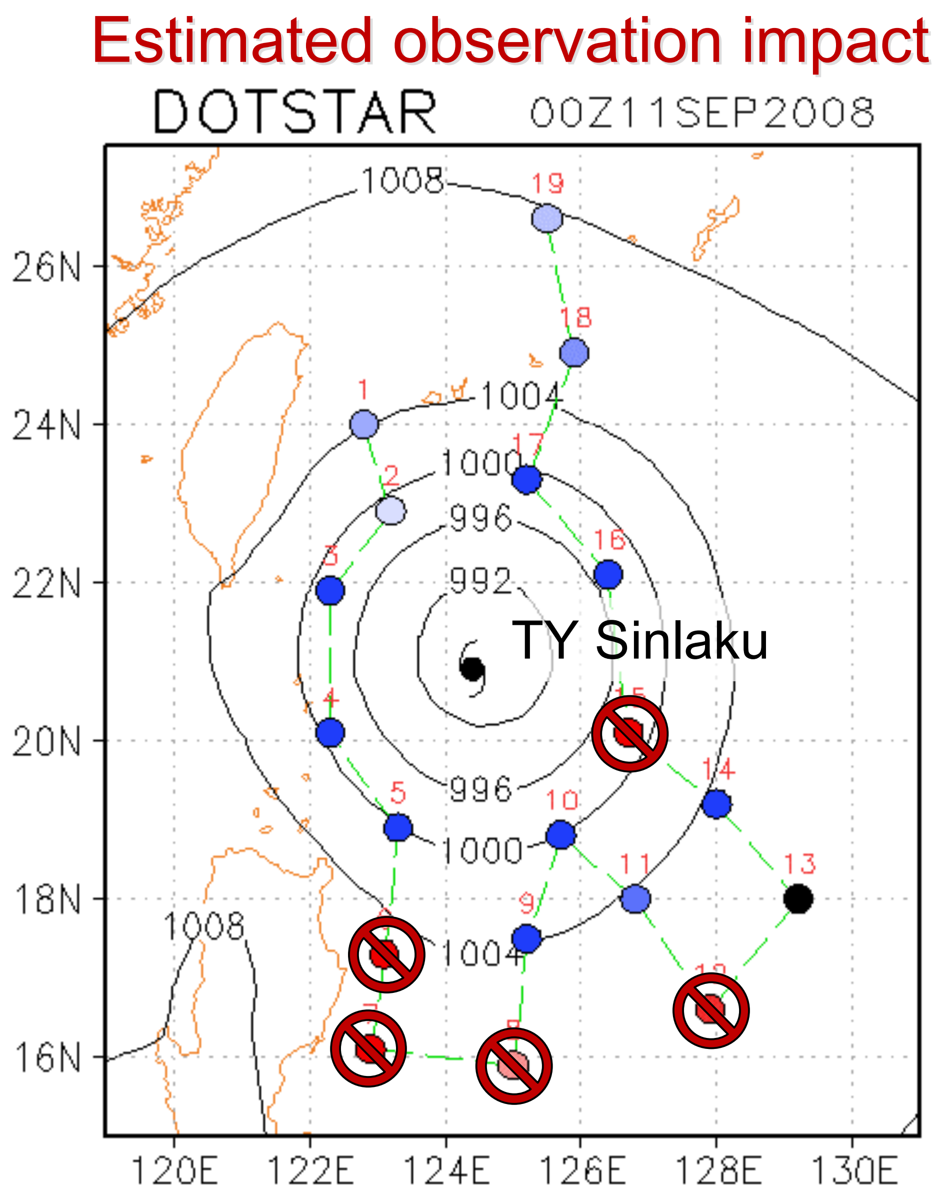
Impact of dropsondes on a Typhoon

Estimated observation impact

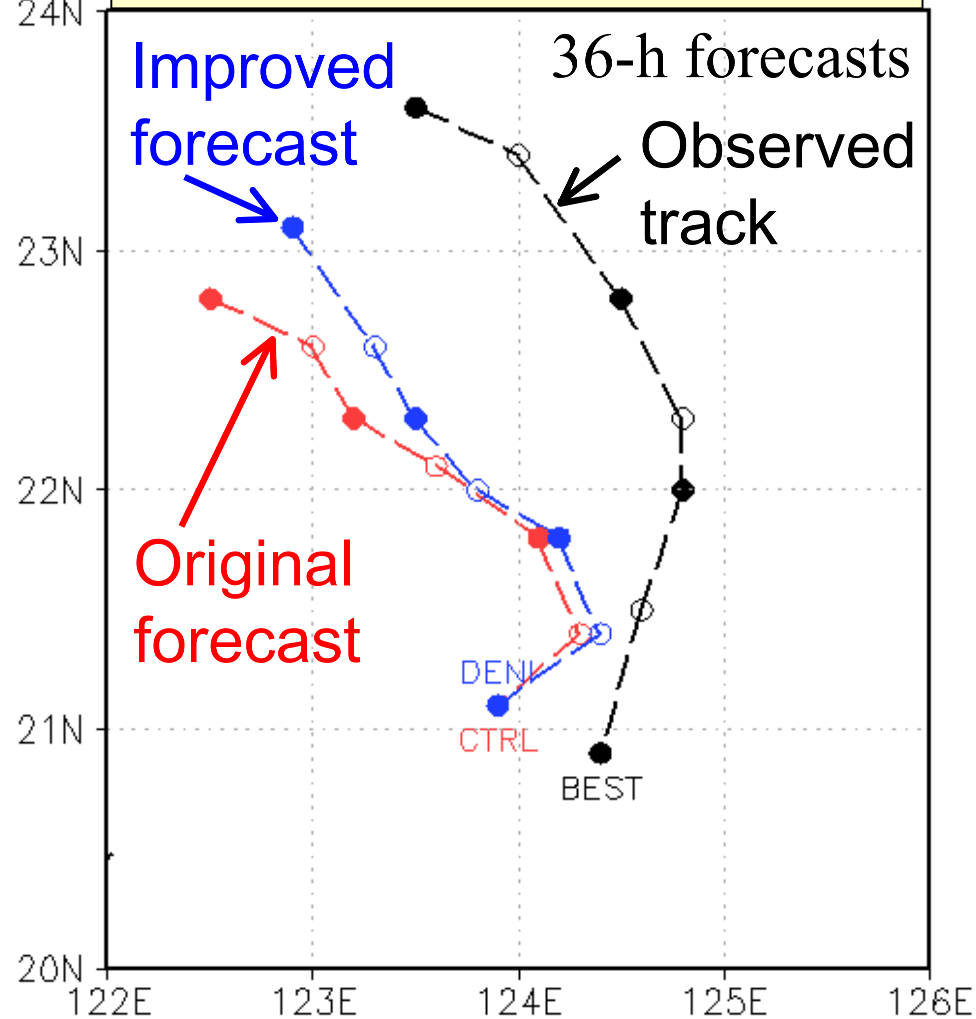
Kunii et al., 2011



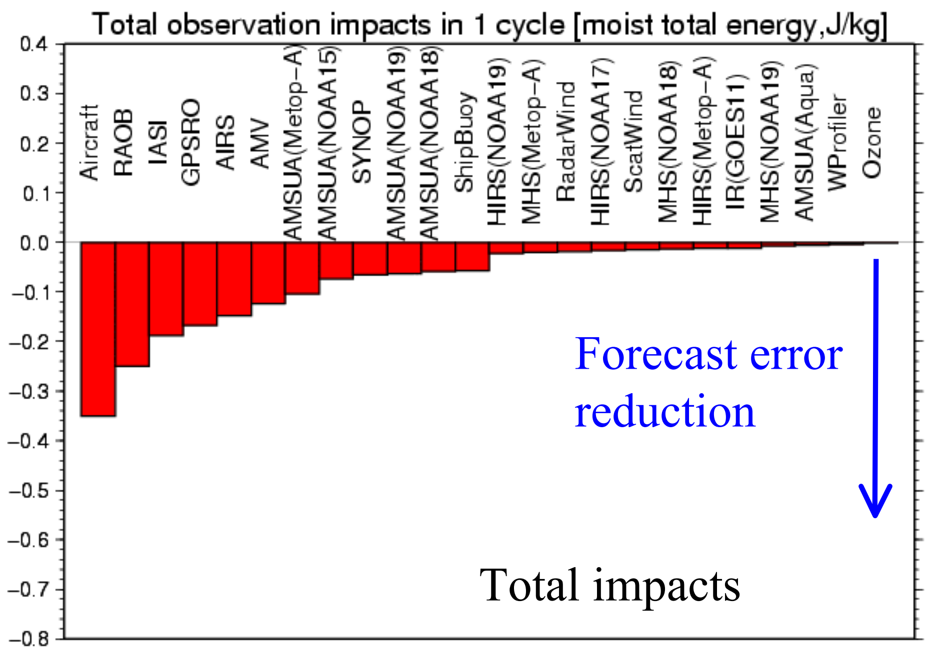
Denying negative impact data improves forecast!

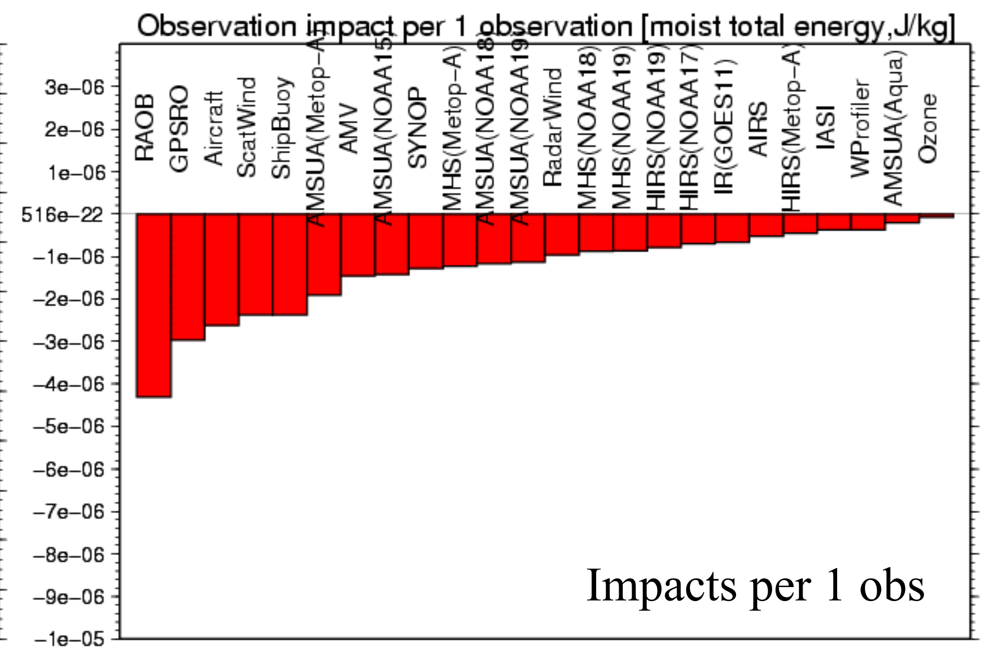


Typhoon track forecast is actually improved!!



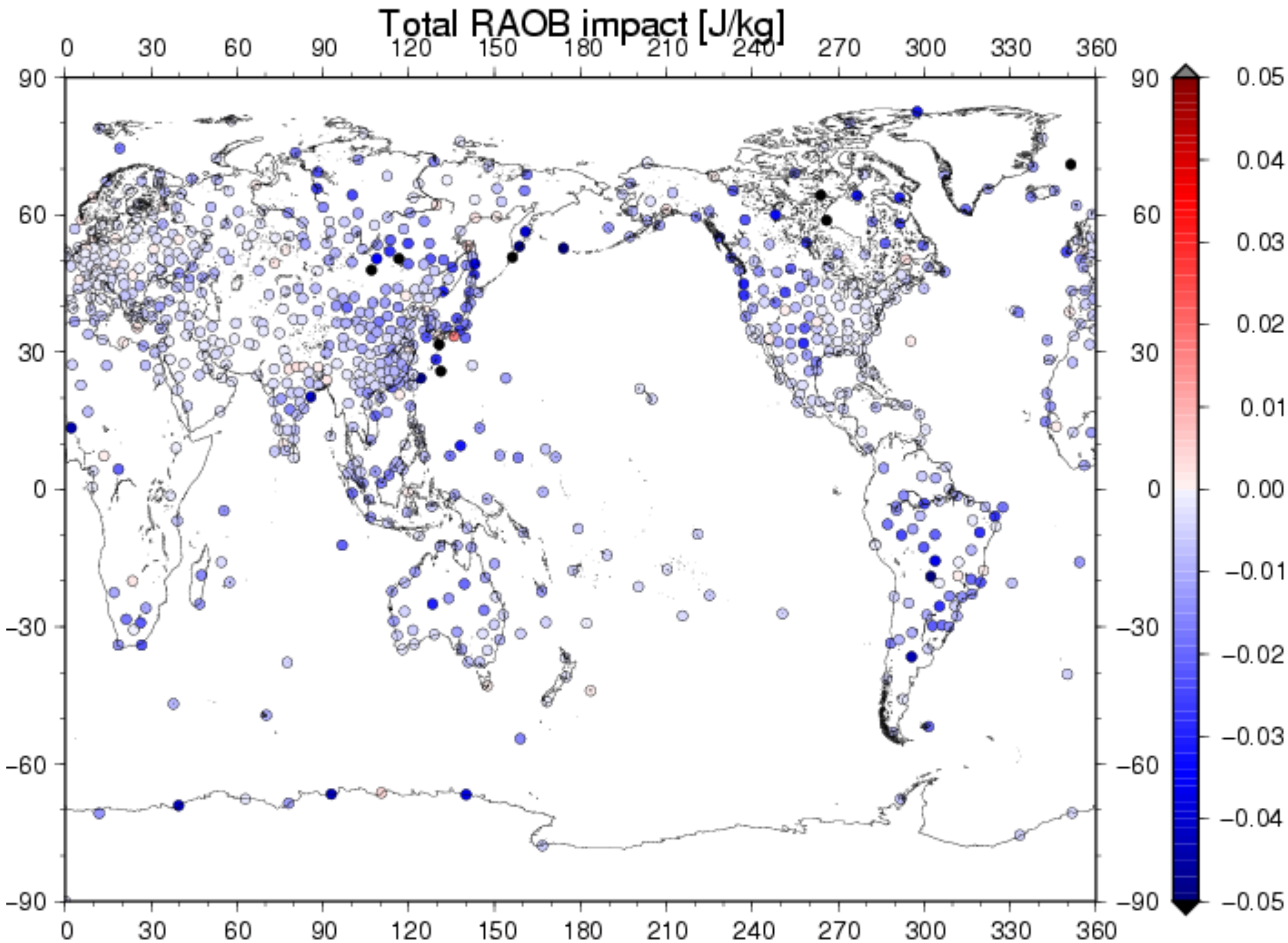
Impact summary: Ota et al., 2012





All observation types have positive impacts on average.
For the total impact, 1: aircraft, 2: AMSU-A, 3: radiosonde, 4: IASI, 5: GPSRO
For impact per 1 obs., 1: radiosonde, 2: GPSRO, 3: aircraft, 4: Scattrometer wind, 5: marine surface observation

Radiosonde impacts: Ota et al. 2012

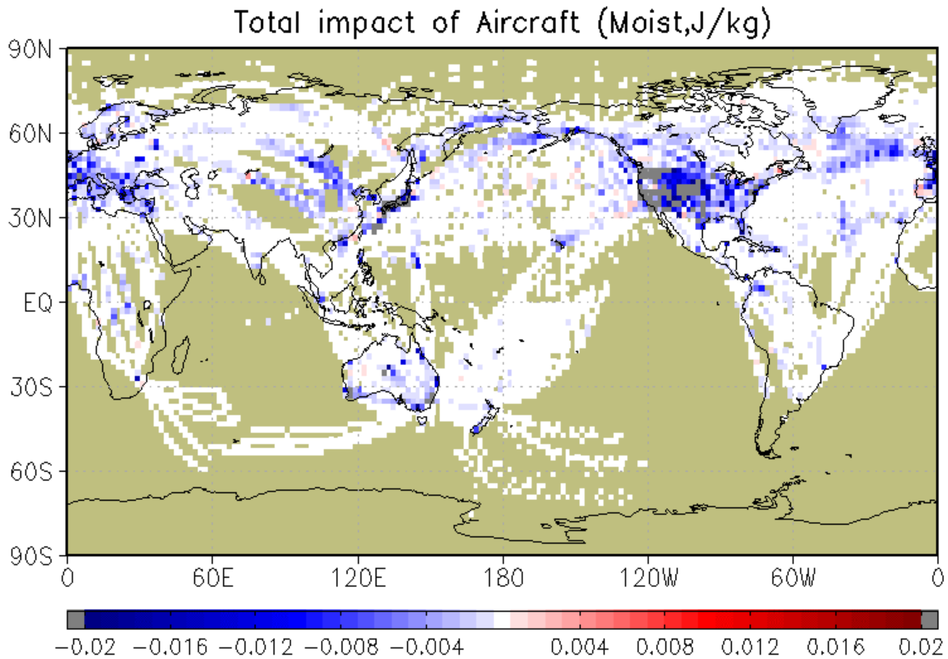


Total impacts of radiosonde (12UTC October 21 to 06UTC October 28)

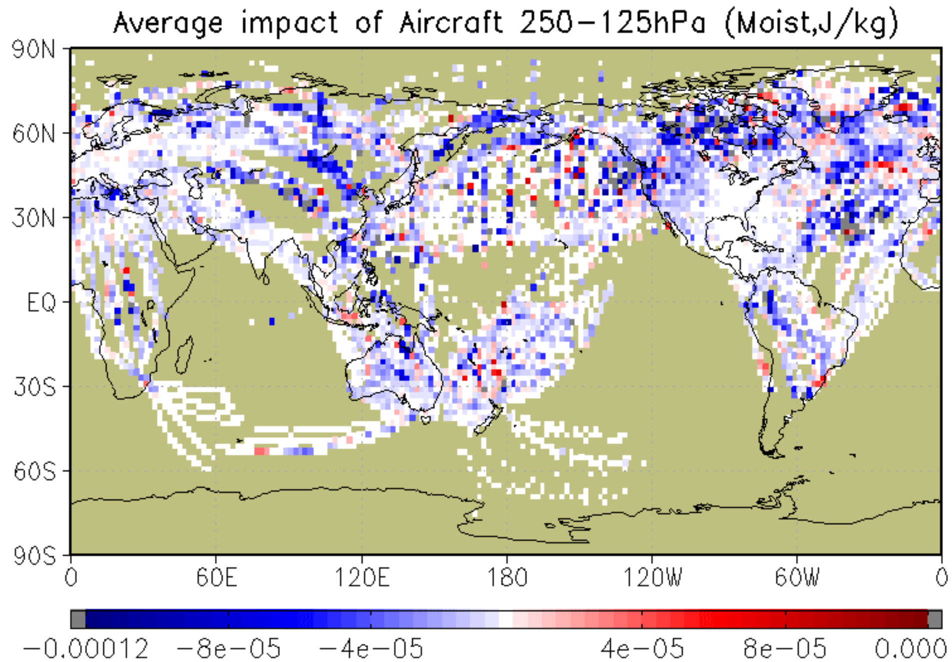
Most observations have positive impacts on average

Relatively large impacts for East Asia, Western US, Canada, and South America.

Aircraft impacts (Ota et al. 2012)



Total impacts, moist total energy



Impacts per 1 obs (250-125 hPa)

Aircraft observations over US, Europe and East Asia have large positive impacts.
The impact of aircraft observations is extremely large over US, however impact per 1 observations is small.

"Carbon Data Assimilation with a Coupled Ensemble Kalman Filter"
Supported by Climate Change Prediction Program in Department of Energy

Simulation mode (SPEEDY)
Develop new methodologies
University of Maryland
Prof. Eugenia Kalnay, Dr. Ji-Sun Kang

Realistic System (CAM/CLM)
Assimilating real observation of GOSAT & AIRS
UC Berkeley
Prof. Inez Fung, Dr. Junjie Liu

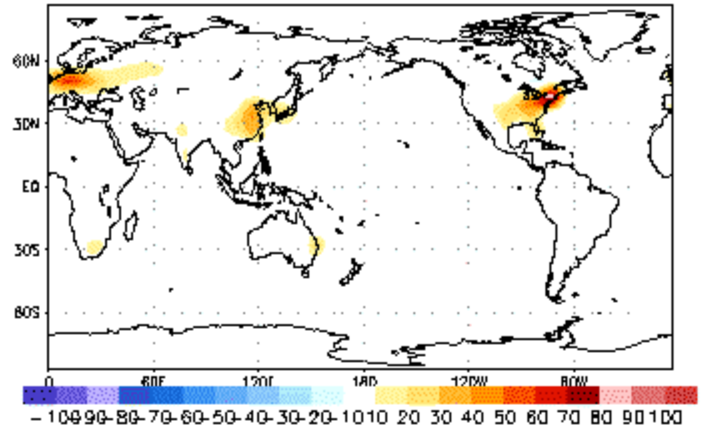
Objectives

- Explore the feasibility of **estimating surface CO_2 fluxes at the model-grid scale** by assimilating atmospheric variables (U, V, T, q, Ps) and atmospheric CO_2 *simultaneously*
  - Consider transport errors in analyzing CO_2 variables
  - No *a-priori* information for CO_2

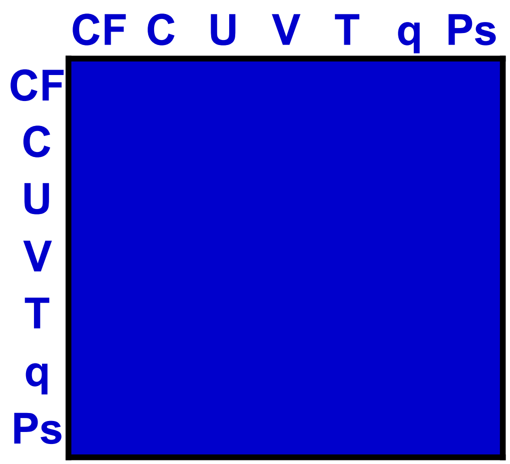
“Localization” of variables

- Analysis of surface CO_2 fluxes assimilating atmospheric CO_2 observations
  - Fossil fuel forcing

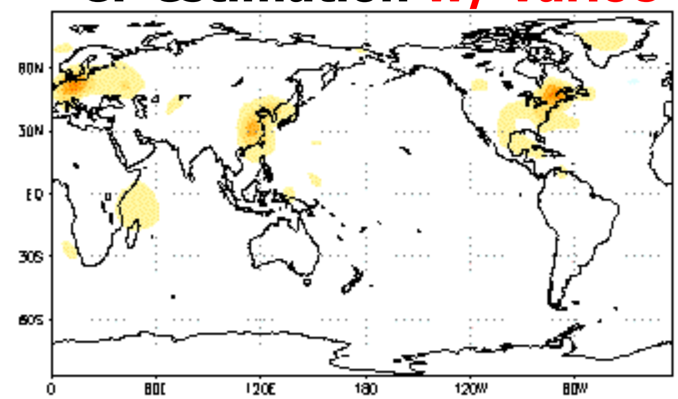
True fluxes



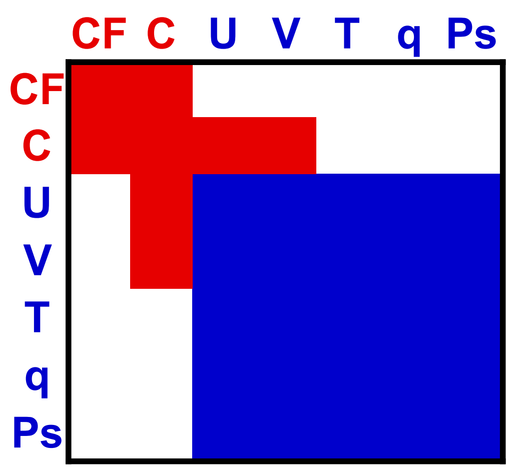
B matrix
Fully multivariate analysis ➔



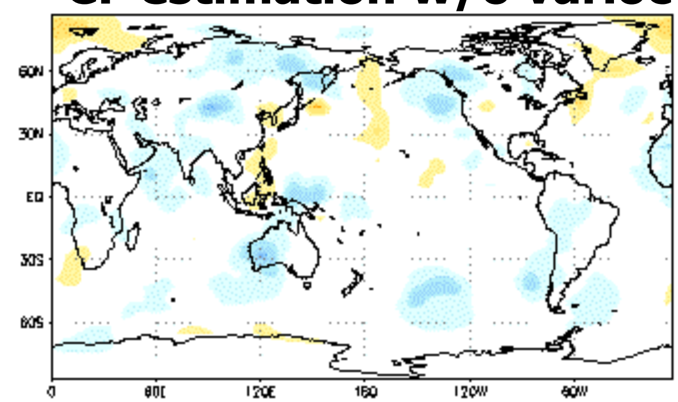
CF estimation w/ varloc



1-way multivariate analysis with variable localization ➔

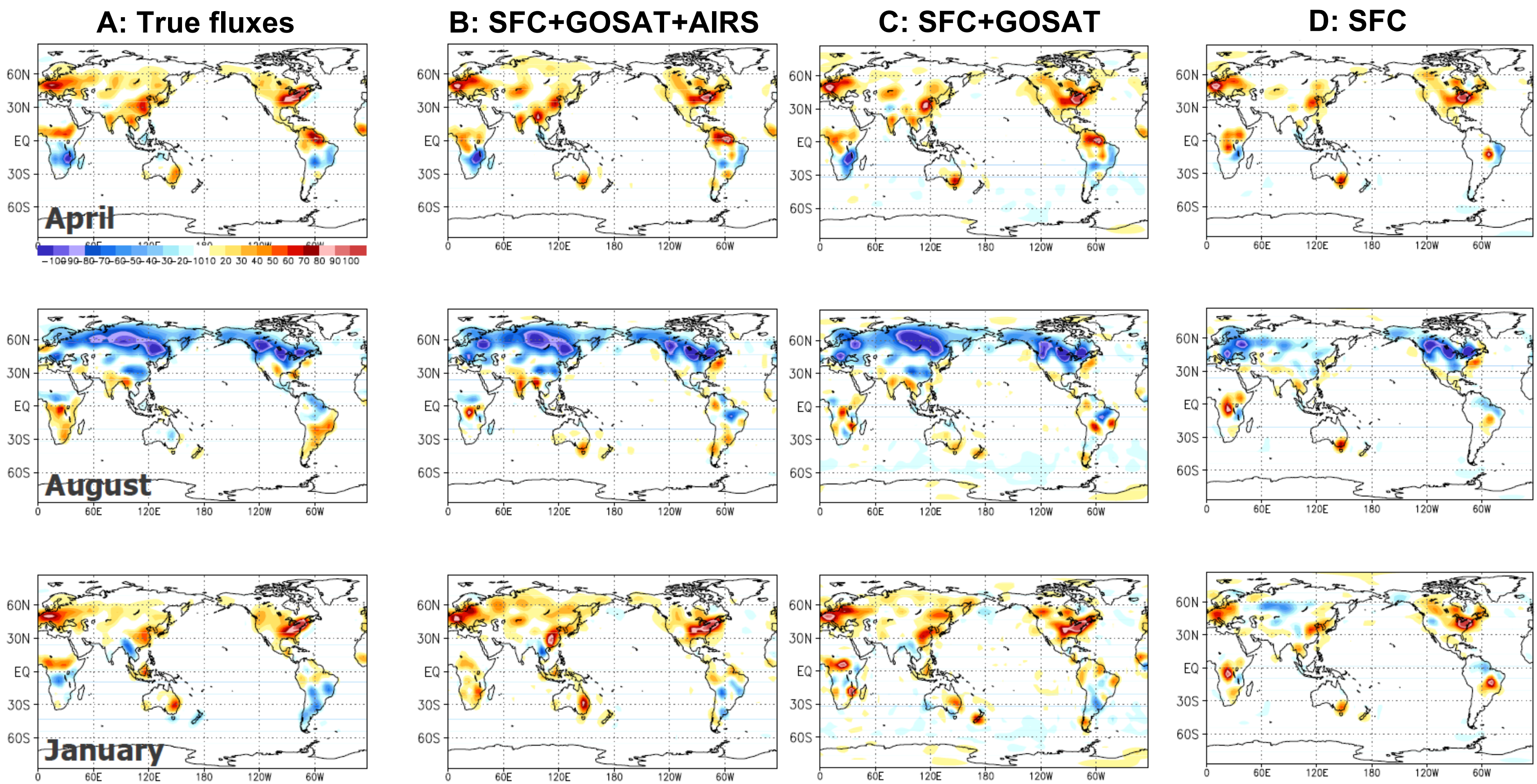


CF estimation w/o varloc



Kang et al. (2011, JGR)

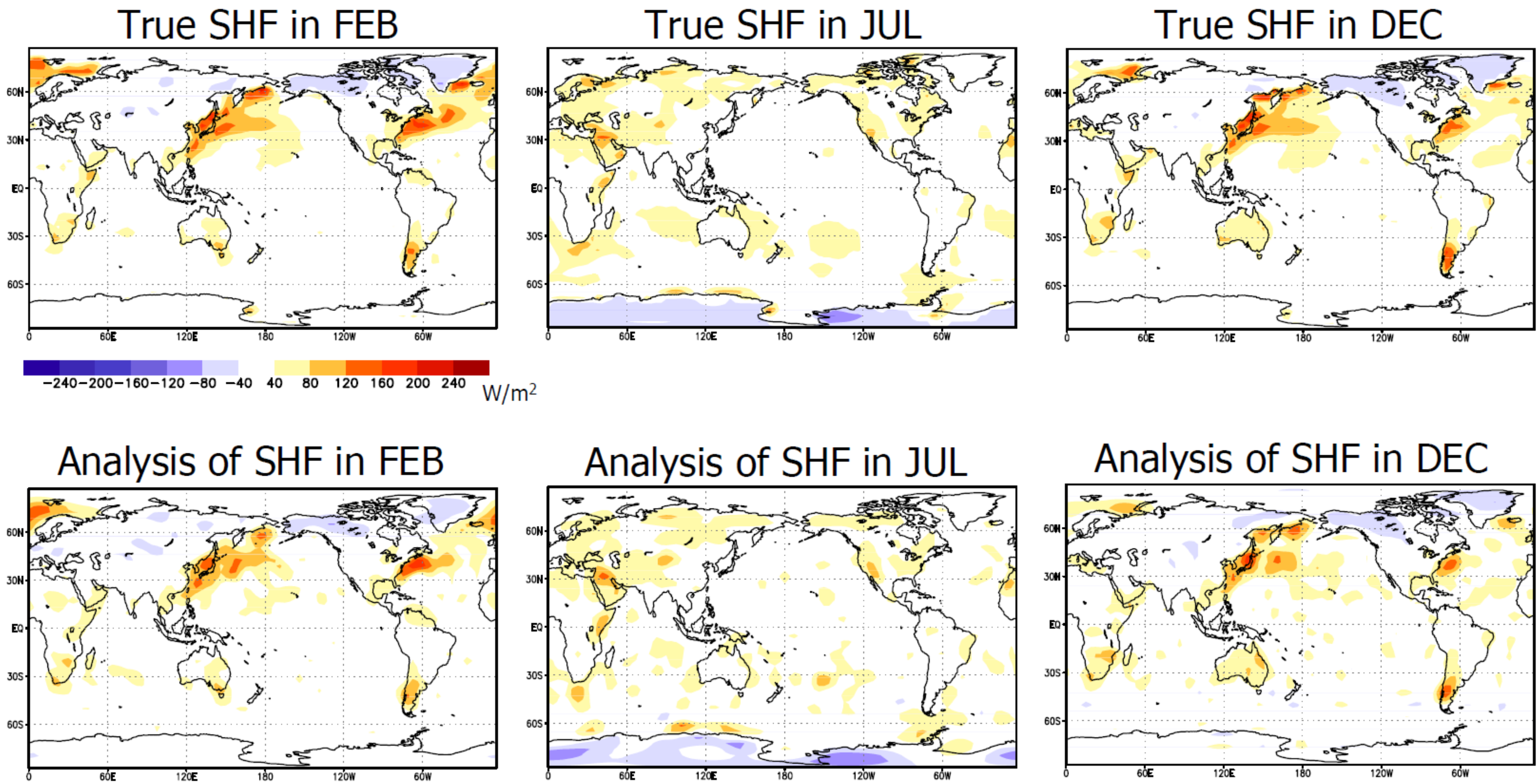
Observation impact



Application to heat/moisture flux estimation

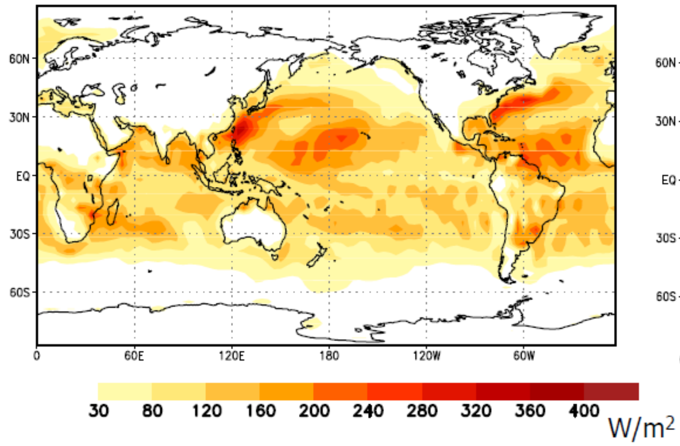
- Can we estimate **surface heat/moisture fluxes** by assimilating atmospheric temperature/moisture observations? ***We can use the same methodology!***
- OSSEs
  - Nature run: SPEEDY
  - Forecast model: SPEEDY with **persistence forecast of sensible/latent heat fluxes (SHF/LHF)**
  - Observations: raob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

Result: Analysis of SHF

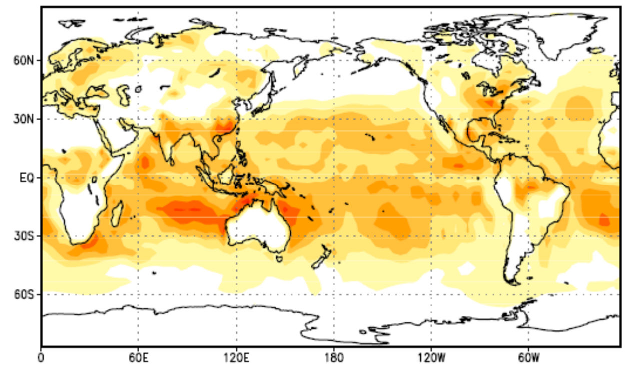


Result: Analysis of LHF

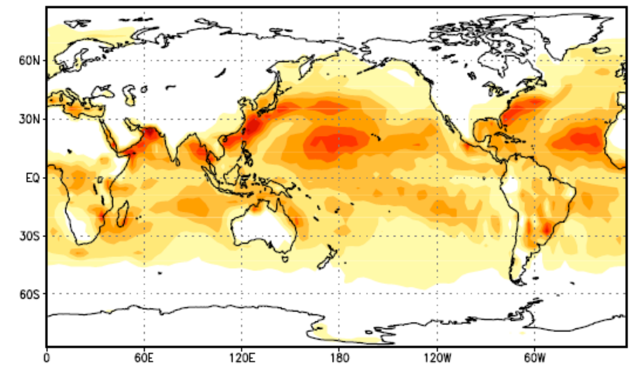
True LHF in FEB



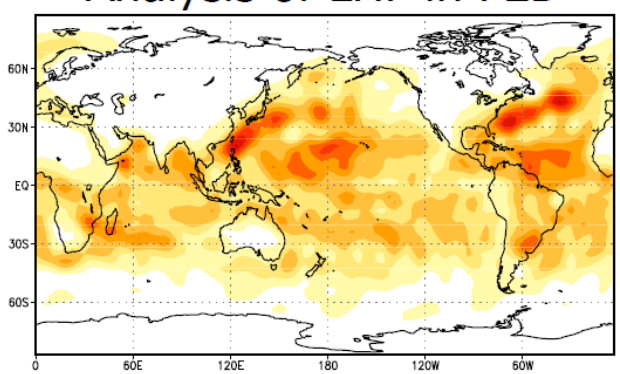
True LHF in JUL



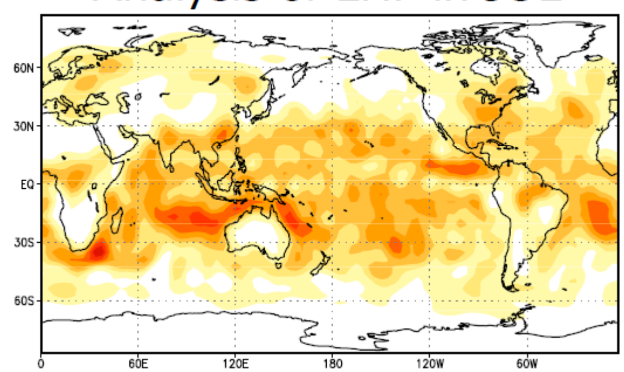
True LHF in DEC



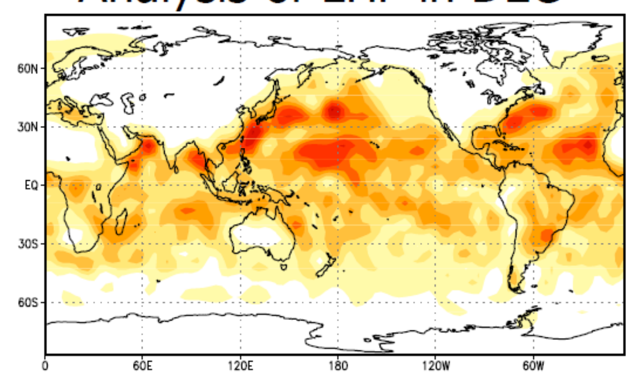
Analysis of LHF in FEB



Analysis of LHF in JUL

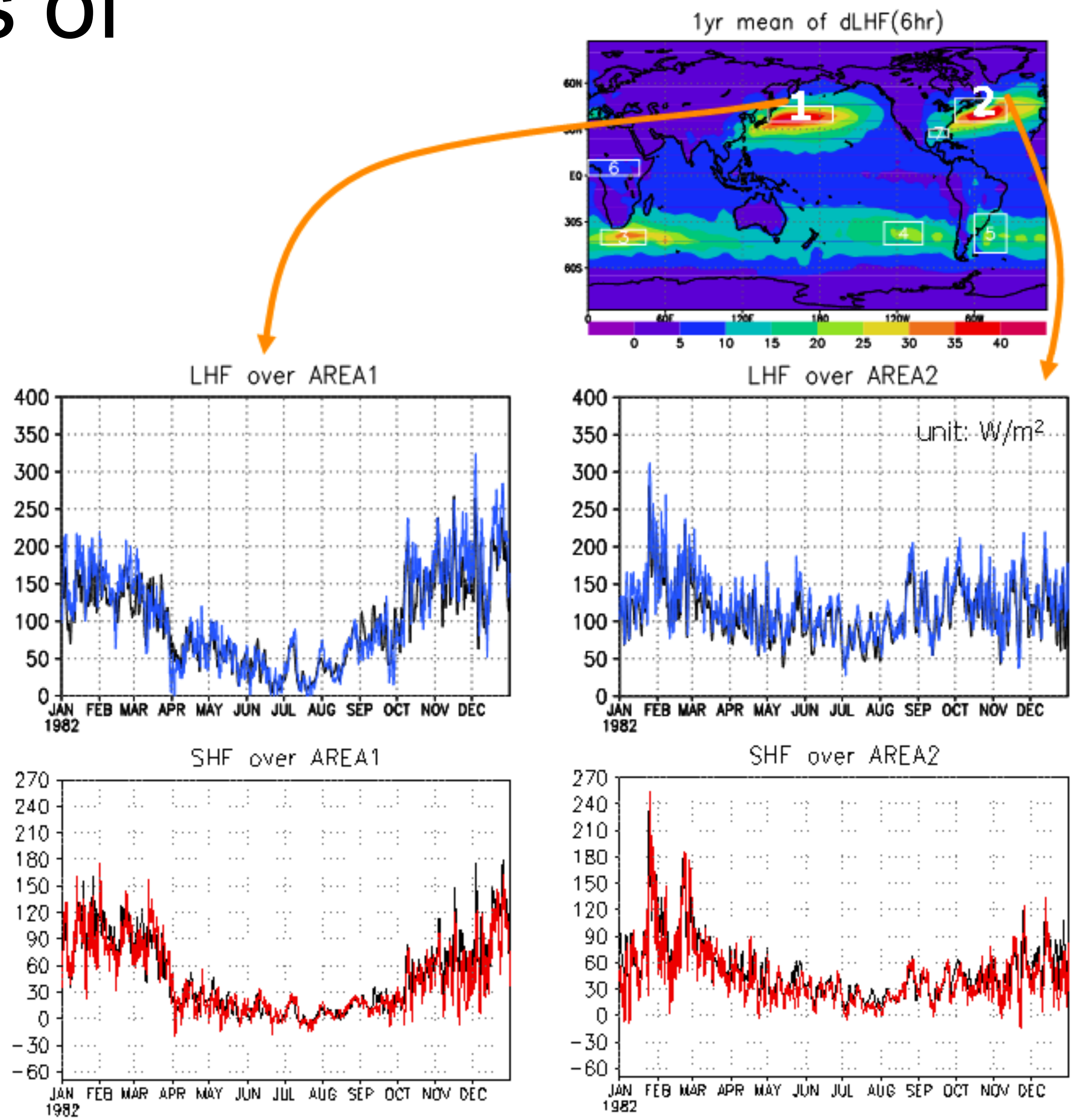


Analysis of LHF in DEC



Time series of LHF/SHF

Recall that LHF & SHF are updated only by the data assimilation here!



Promising results from the estimation of "evolving parameters" with data assimilation

Tamara Singleton's thesis:

Data Assimilation Experiments with a Simple Coupled Ocean-Atmosphere Model

Questions explored:

-- Which one 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?

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	$c, c_z = 1$ $c_e = 0.08$
τ	time scale	$\tau = 0.1$
σ, b, and r	Lorenz parameters	$\sigma=10$, $b=8/3$, and $r=28$
k_1, k_2	Uncentering parameters	$k_1=10$ $k_2 = -11$

Extratropical atmosphere

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

Tropical atmosphere

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

Ocean

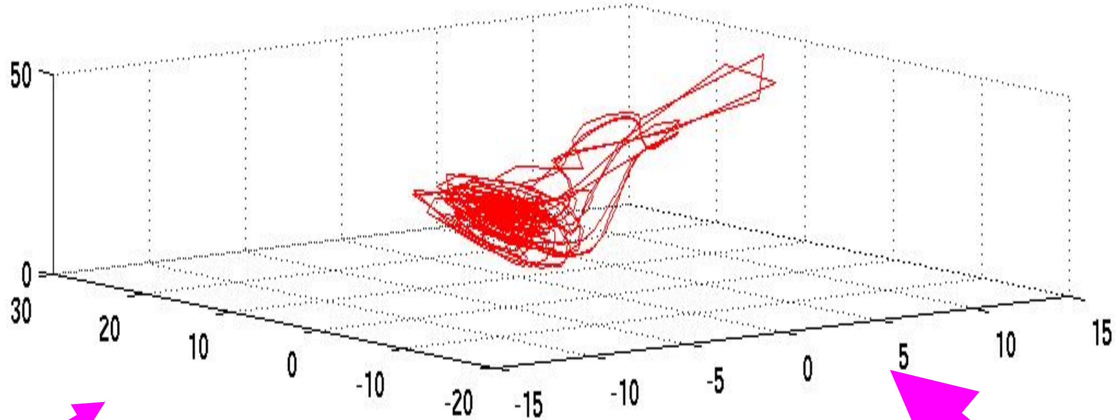
$$\dot{X} = \tau\sigma(Y - X) - c(x_t + k_2)$$
$$\dot{Y} = \tau rX - \tau Y - \tau XZ + c(y_t + k_2)$$
$$\dot{Z} = \tau XY - \tau bZ + c_z z_t$$

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

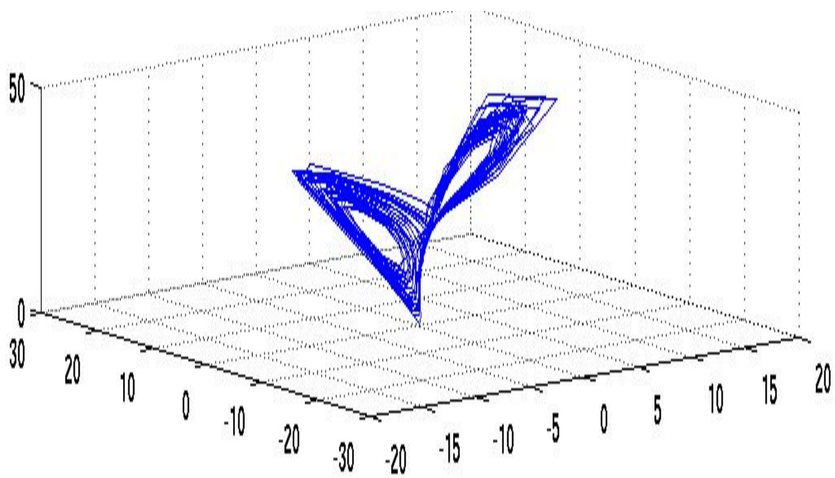
Simple Coupled Ocean-Atmosphere Model (Peña and Kalnay, 2004)

Coupling strength

Tropical Atmosphere

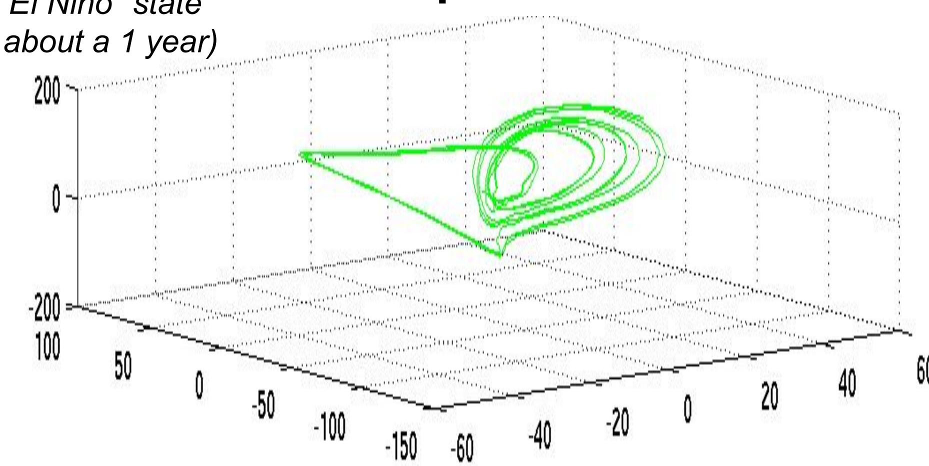


Extra-tropical Atmosphere

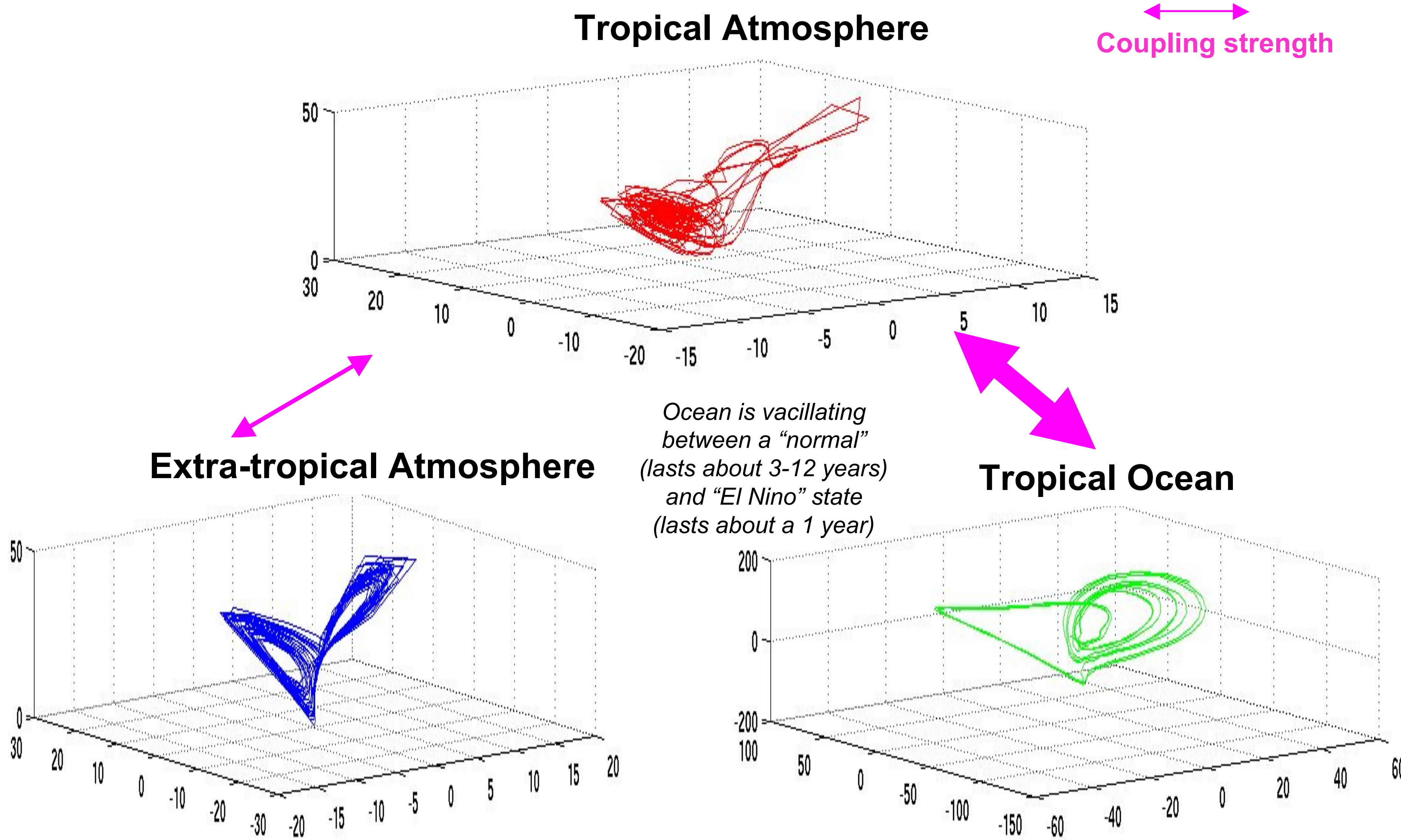


Ocean is vacillating between a "normal" (lasts about 3-12 years) and "El Nino" state (lasts about a 1 year)

Tropical Ocean

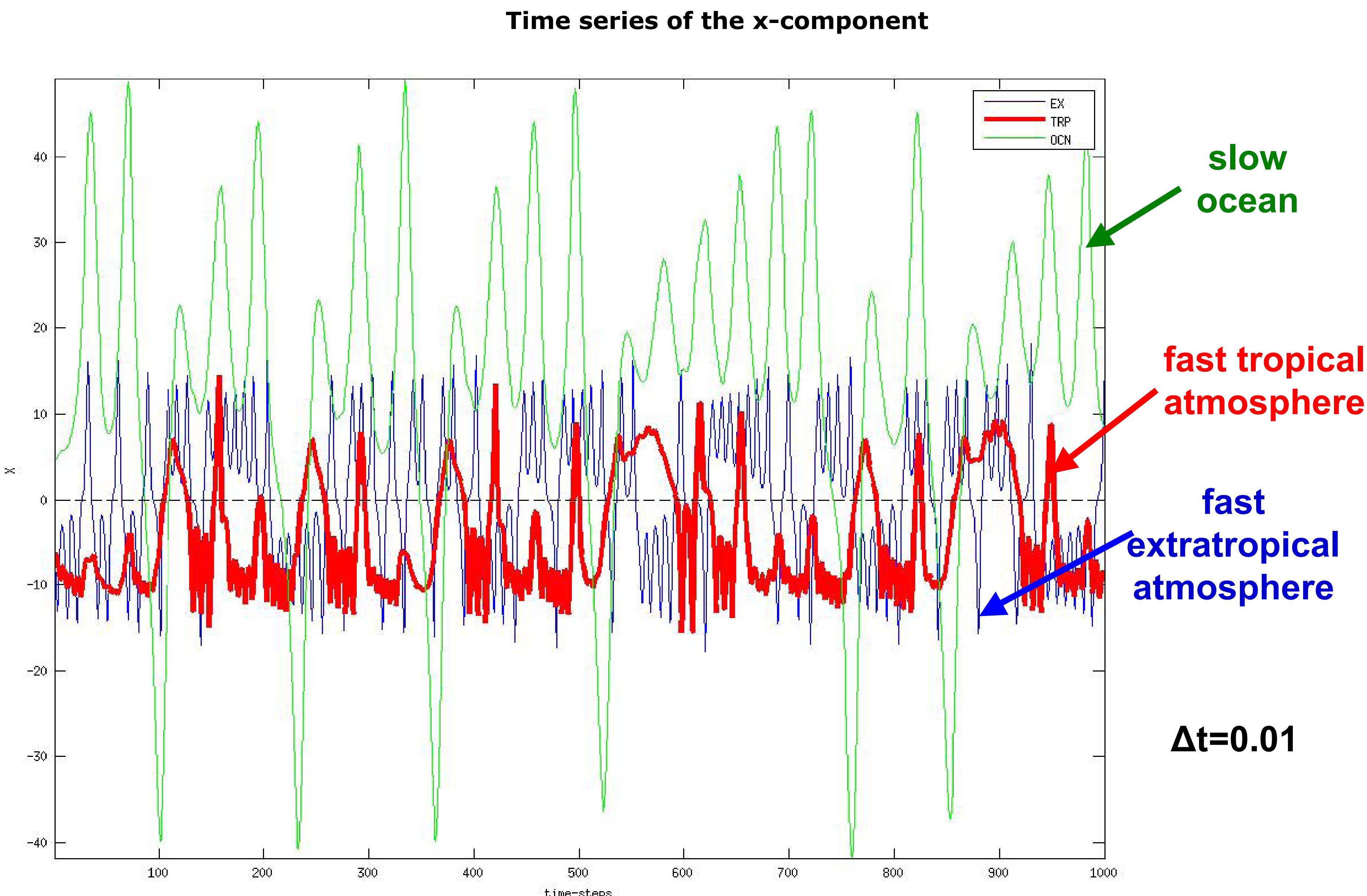


Simple Coupled Ocean-Atmosphere Model (Peña and Kalnay, 2004)



We compare 4D-Var and EnKF with this simple coupled model

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



We compare 4D-Var and EnKF with this simple coupled model

Data Assimilation Experiment Design

- **Simple Coupled Ocean-Atmosphere Model (perfect model)**
  - Used to create the “true” trajectory
- **Observations**
  - Generated from the nature run plus “random errors” with 1.41 s.d.
  - Every 8 time steps of a simulation
- **Perform coupled and uncoupled ocean data assimilations with several EnKF, 4D-Var, and ECCO-4D-Var**

- **Compute RMS errors** of the difference between the analysis and the true solution.

- **Lengthen assimilation windows, from 8 to 320 steps**

- **Perform fully coupled data assimilation (ETKF, 4D-Var), and just ocean assimilation (LETKF, 4D-Var and ECCO)**

EnKF-Based Methods

Description of EnKF-based methods

Method	Assimilating	Observations	Special Features
ETKF (Fully coupled)	Fast and slow variables simultaneously	Available at the end of a window (analysis time)	
4D-ETKF (Fully coupled)	Fast and slow variables simultaneously	Available throughout an assimilation window	4-dimensional
ETKF-QOL (Fully coupled)	Fast and slow variables simultaneously	Available at analysis time	Uses quasi-outer loop to improve the initial analysis mean
LETKF (Separate Ocean)	Fast and slow variables separately	Available at analysis time	Subsystem localization
4D-LETKF (Separate Ocean)	Fast and slow variables separately	Atmos: Available at analysis time Ocean: Available throughout an assimilation window	4-dimensional Subsystem localization

Variational Data Assimilation Experiments:

Fully Coupled 4D-Var

Ocean only 4D-Var

ECCO-like Ocean 4D-Var

Fully coupled 4D-Var : the Cost Function

- In 4D-Var, a cost function is minimized to produce an optimal analysis.
  - The cost function measures the distance between the model with respect to the observations and with respect to the background state.

- The analysis is obtained by minimizing the cost function given by

$$J(\mathbf{x}_{t_0}) = \frac{1}{2}\left[\mathbf{x}_{t_0} - \mathbf{x}^b_{t_0}\right]^T B_0^{-1}\left[\mathbf{x}_{t_0} - \mathbf{x}^b_{t_0}\right] + \frac{1}{2}\sum_{i=1}^{N}\left[H(\mathbf{x}_{t_i}) - \mathbf{y}^o_{t_i}\right]^T \mathbf{R}_{t_i}^{-1}\left[H(\mathbf{x}_{t_i}) - \mathbf{y}^o_{t_i}\right]$$

J^b - "background" cost function

J^o- "observation" cost function

where the control variables are

$$\mathbf{x}_0 = \left(x_e^0, y_e^0, z_e^0, x_t^0, y_t^0, z_t^0, X^0, Y^0, Z^0\right)^T$$

Initial model state for extratropical atmos

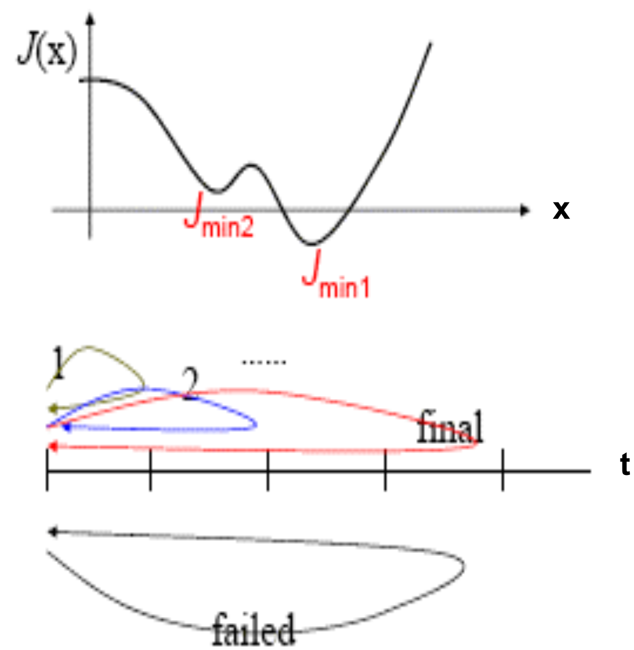
Initial model state for tropical atmos.

Initial model state for ocean

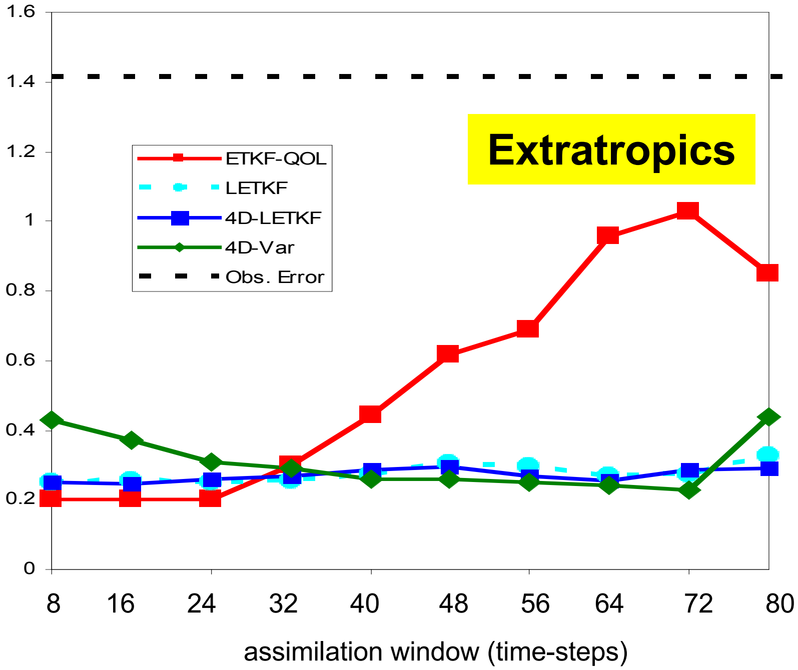
4D-Var: Quasi-static Variational Data Assimilation (QVA)

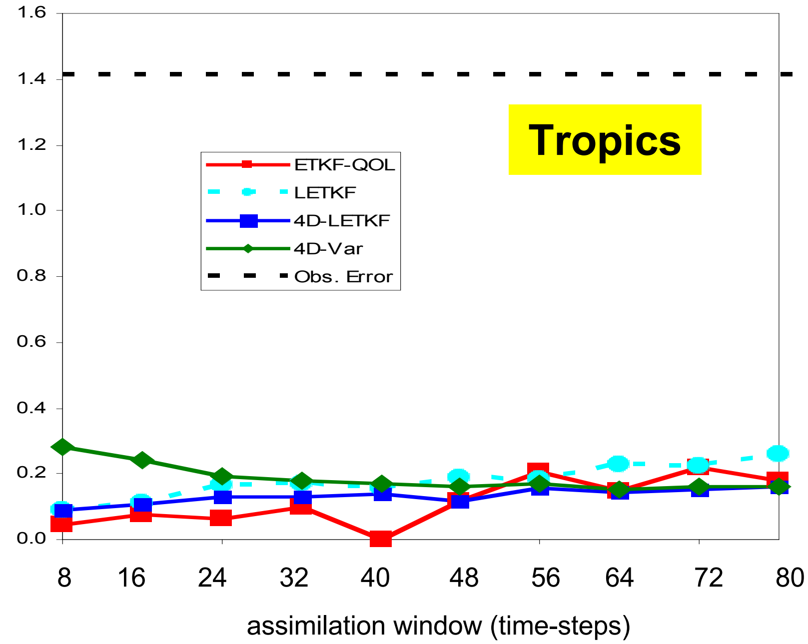
- For longer windows, multiple minima are a problem for 4D-Var minimization (Pires et al., 1996).
- Also for longer assimilation windows, non-Gaussian perturbations of the observation error and background error -> in non-quadratic cost functions
- Pires et al. (1996) proposed the Quasi-static Variational Data Assimilation (QVA) approach.

Schematic of multiple minima and increasing window size (Pires et al, 1996)

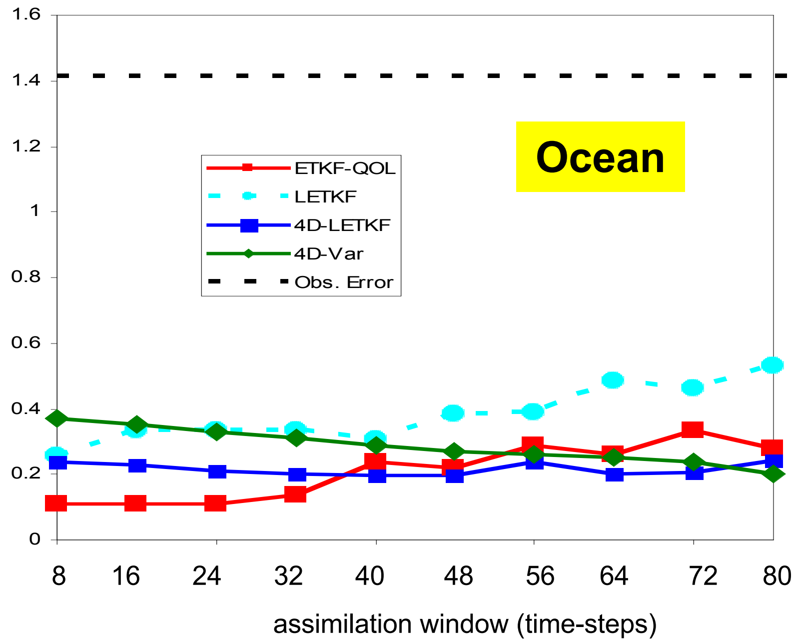


Fully coupled 4D-Var and EnKF: shorter windows



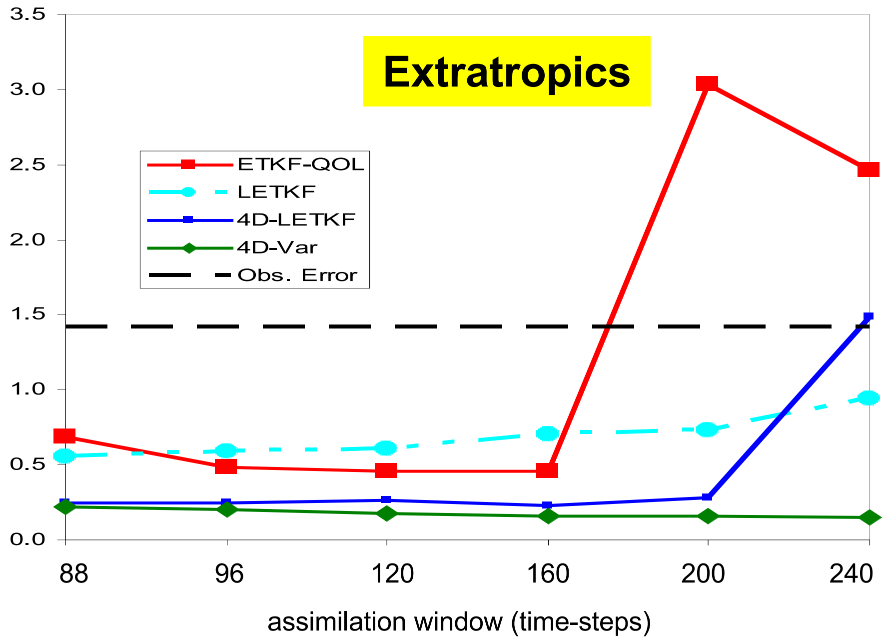


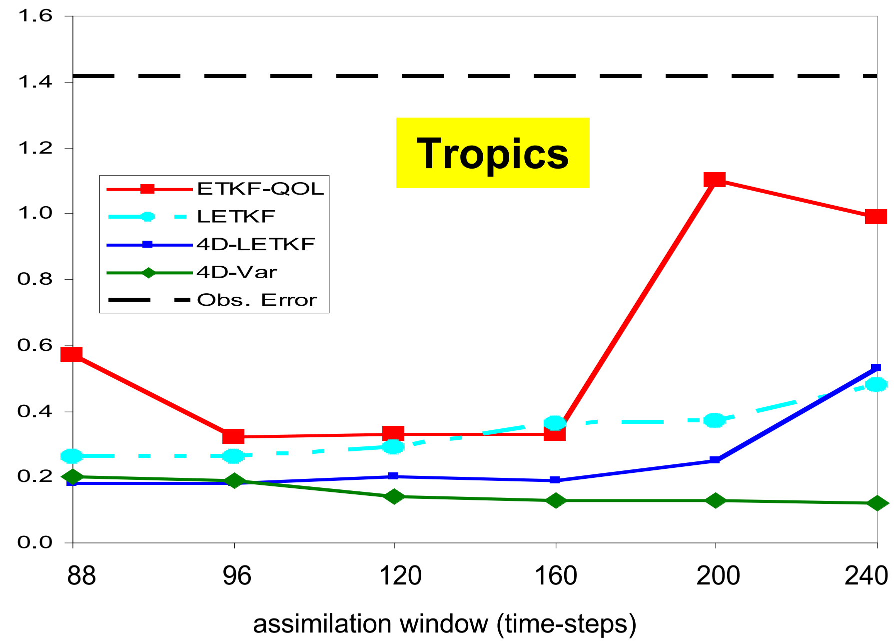
ETKF-QOL provides the best analysis for very short windows



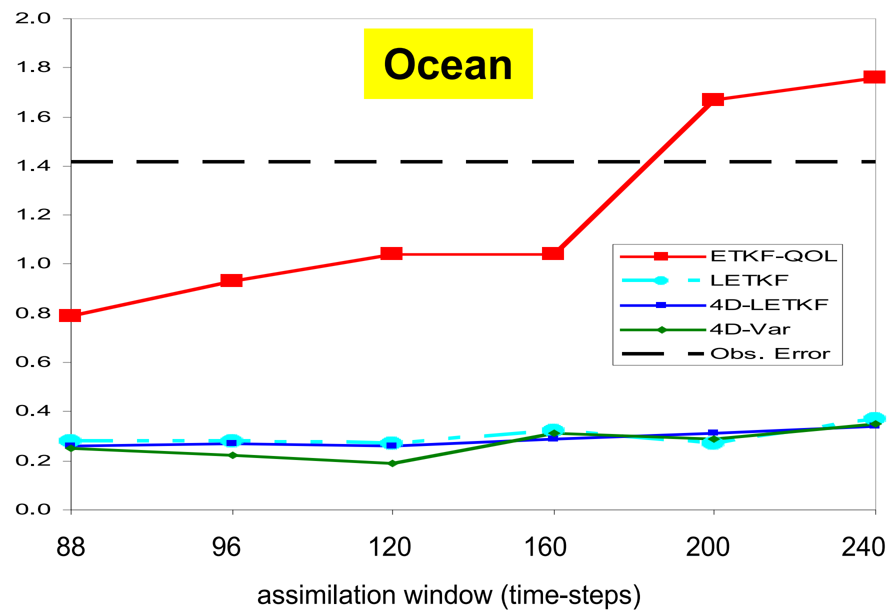
4D-Var competes with EnKF-based methods for longer windows

Fully coupled 4D-Var vs. EnKF: longer windows





Coupled 4D-Var and EnKF competitive for longer windows



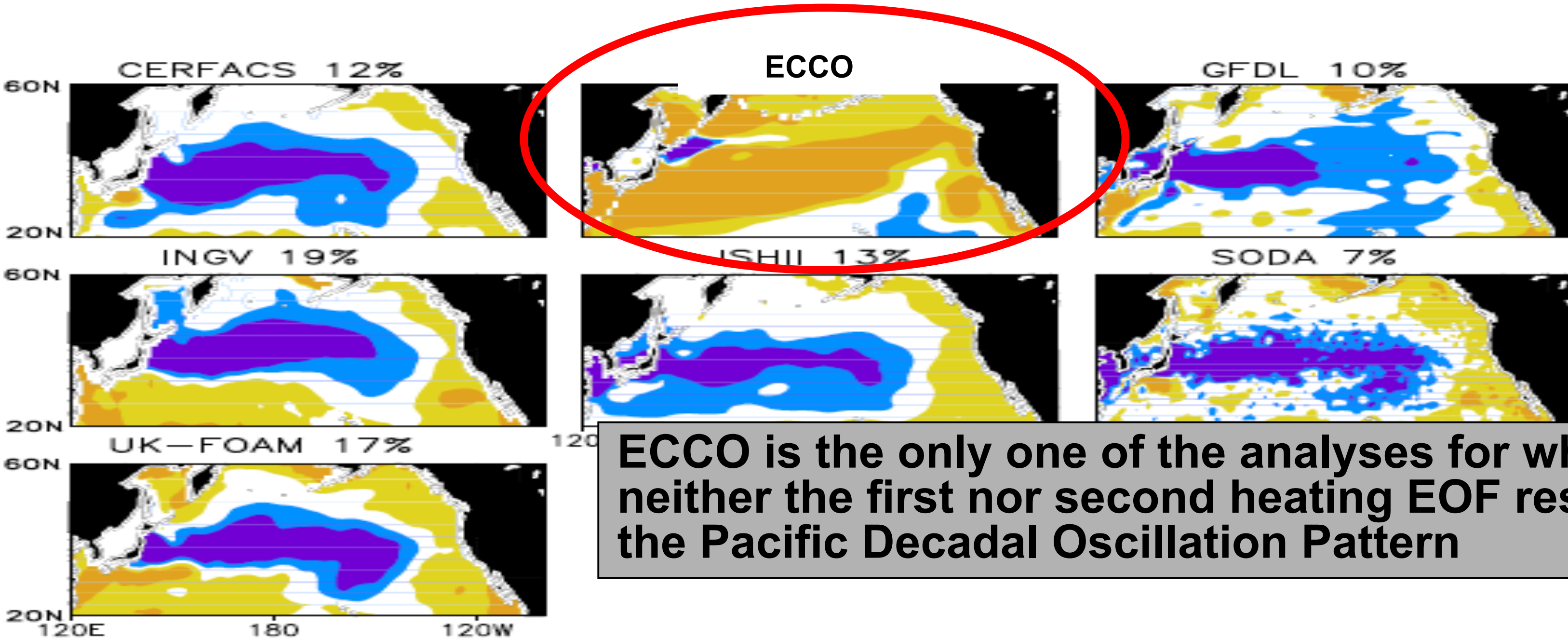
Fully coupled 4D-Var vs EnKF summary

- **We developed fully coupled 4D-Var and EnKF systems for the simple coupled ocean-atmosphere model**
- **4D-Var needs tuning the amplitude of the background error covariance B. EnKF needs tuning of inflation (or adaptive inflation).**
- **Lengthening the assimilation windows and applying QVA improves the 4D-Var analysis because 4D-Var "forgets" B. But longer windows are <u>more</u> expensive…**
- **Fully coupled EnKF are optimal for short windows. Short windows are <u>less</u> expensive…**
- **The optimal configurations (short windows for EnKF and long windows for 4D-Var) have similar accuracy.**

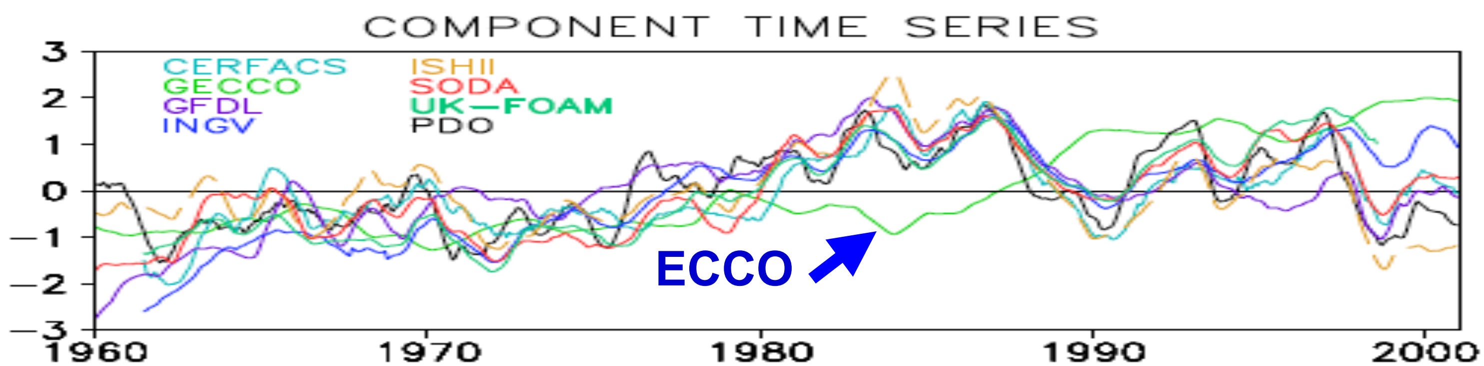
ECCO-like 4D-Var

- 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

Motivation: Comparison of Ocean Analyses



ECCO is the only one of the analyses for which neither the first nor second heating EOF resemble the Pacific Decadal Oscillation Pattern



Carton and Santorelli (2008) plot of the First Empirical Orthogonal Eigenfunction of monthly heat content anomaly in the latitude band 20N-60N Explained variance is shown on the title line. Lower panel shows the corresponding component time series annually averaged along with the Pacific Decadal Oscillation Index of *Mantua et al. (1997)* in black.

ECCO-like 4D-Var: Cost Function includes all surface fluxes as control variables

$$J = \frac{1}{2}[\mathbf{x}_{0,f} - \mathbf{x}^{b,nfe}]^T (\mathbf{B}^{0,nfe})^{-1} [\mathbf{x}_{0,f} - \mathbf{x}^{b,nfe}] + \frac{1}{2}\sum_{i=1}^{N} [H\mathbf{x}_{t_i} - \mathbf{y}^o_{t_i}]^T (\mathbf{R}^{-1}_{t_i}) [H\mathbf{x}_{t_i} - \mathbf{y}^o_{t_i}]$$

$$\mathbf{B}^{\mathbf{0,nfe}} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$$

where

$$\mathbf{x}_{0,f} = \begin{pmatrix} X_0, Y_0, Z_0 & f_1^1, f_2^1, f_3^1 & f_1^2, f_2^2, f_3^2 & \ldots & f_1^n, f_1^n, f_1^n \end{pmatrix}^T$$

Initial model state

Fluxes for first 8 time steps

Fluxes for last 8 time steps

$$\mathbf{x}^{b,nfe} = \begin{pmatrix} X^b, Y^b, Z^b & f_1^{nfe,1}, f_2^{nfe,1}, f_3^{nfe,1} & f_1^{nfe,2}, f_2^{nfe,2}, f_3^{nfe2} & \ldots & f_1^{nfe,n}, f_1^{nfe,n}, f_1^{nfe,n} \end{pmatrix}^T$$

Background state for the ocean

NCEP-like flux estimates for last 8 time steps

NCEP-like flux estimates for first 8 time steps

ECCO-like 4D-Var: Experimental Design

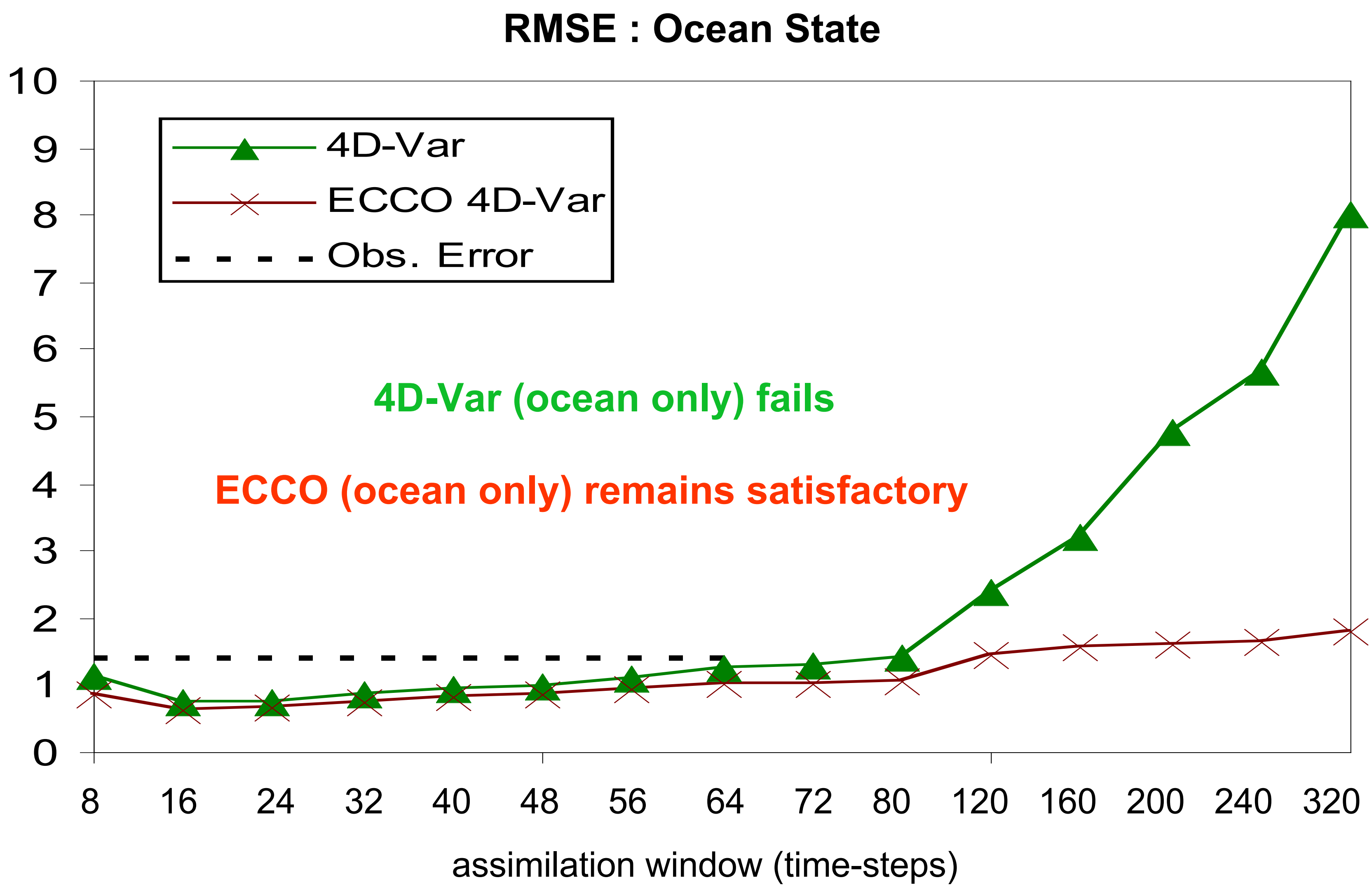
- **Observations**
  - Same as EnKF and 4D-Var experiments
- **Forecast Model**
  - Slow subsystem of coupled model with fluxes changing after every 8 time steps
- **Data Assimilation: ECCO Ocean 4D-Var**
  - **Control variables are <u>initial ocean state and flux terms</u>**
  - Prescribed background error covariance from NMC method
  - Varied length of assimilation windows: 8 – 320 time steps
- **Comparison with: Ocean 4D-Var**
  - **Control variables are <u>initial ocean state</u>**
  - Prescribed background error covariance from NMC method
  - Varied length of assimilation windows: 8 – 320 time steps

Comparison of ECCO-like & Ocean 4D-Var

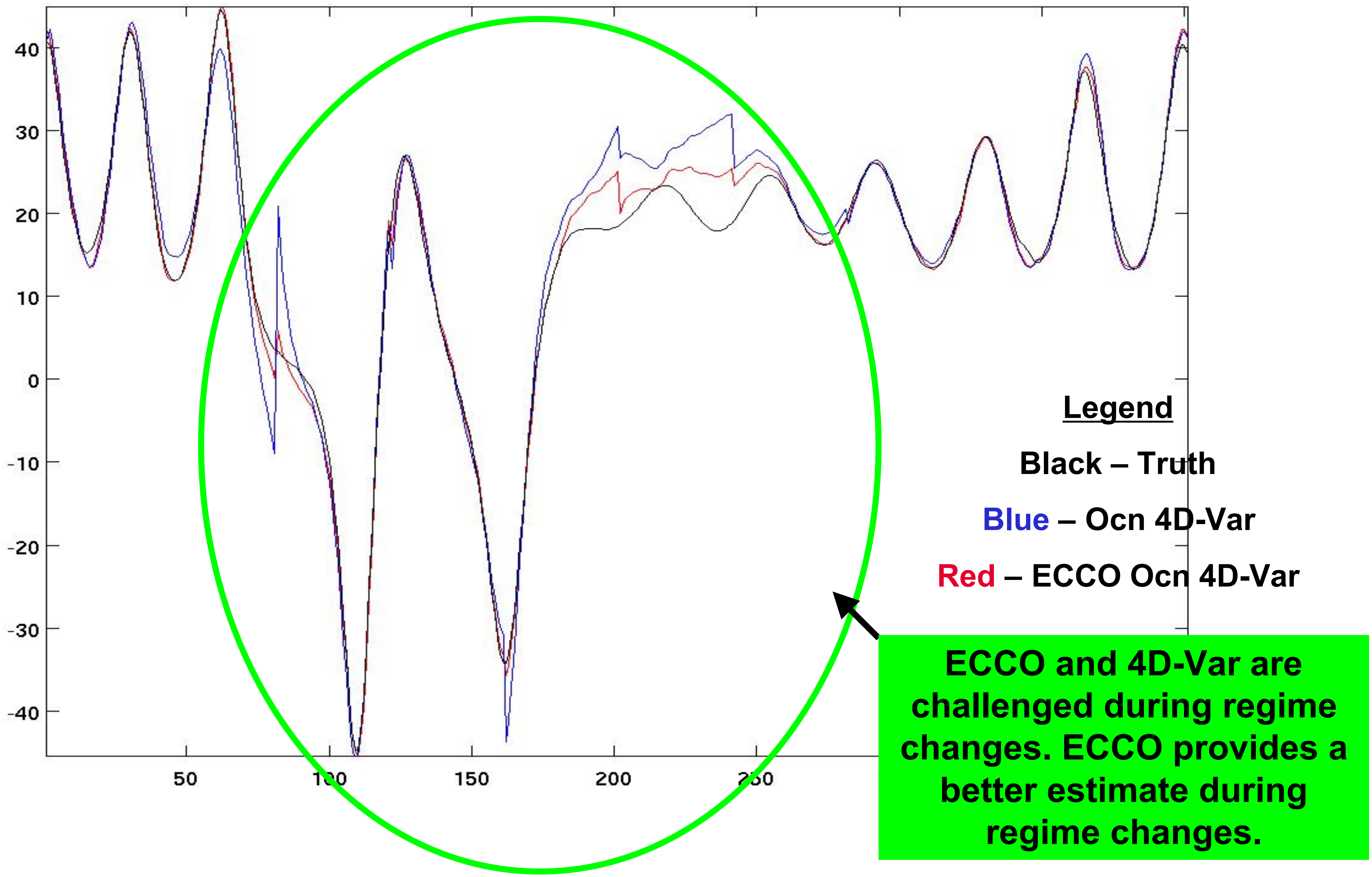
QVA APPLIED

OCEAN ONLY

Obs. s.d. error = 1.41 for ocean

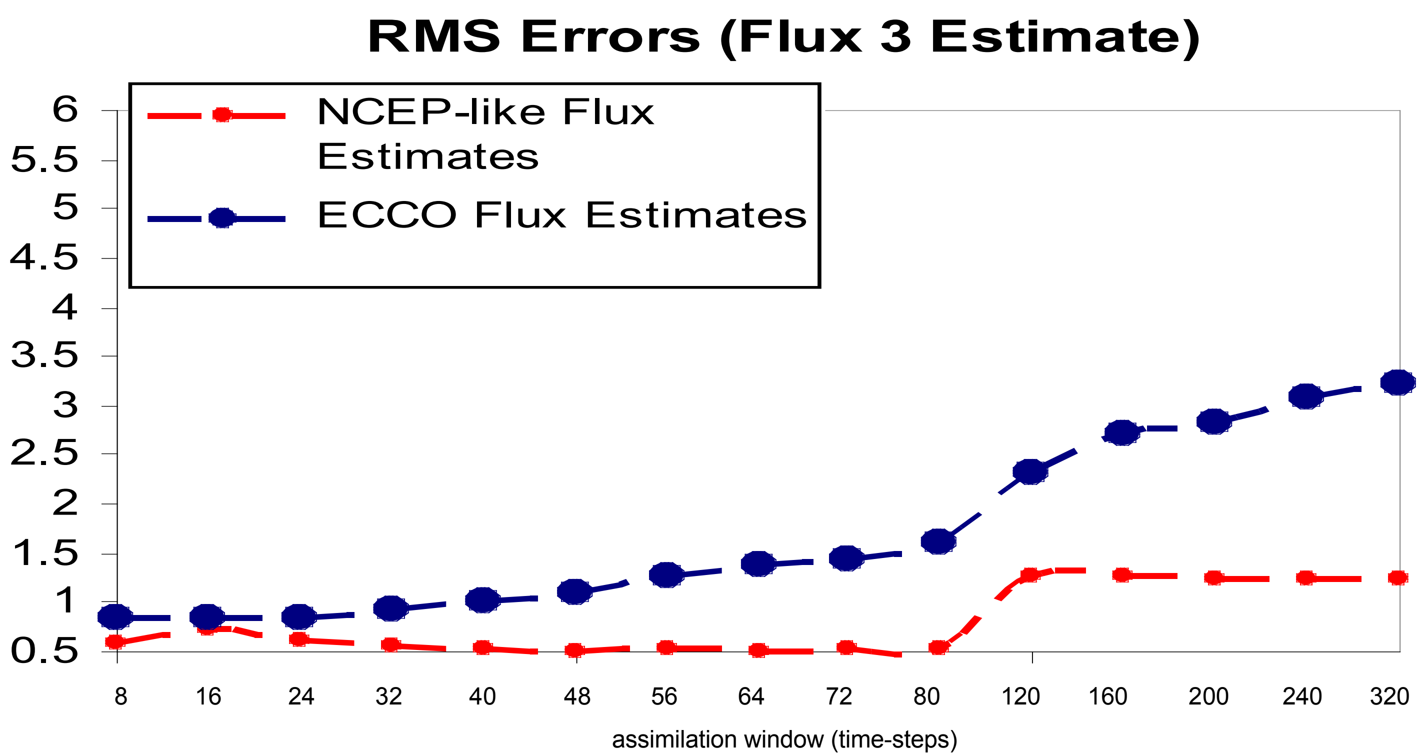


By using sfc fluxes as control variables, ECCO can use very long windows



This is because in ECCO the fluxes "adapt" in order to improve the analysis!

Are the ECCO fluxes more accurate?



ECCO does **not** improve the flux estimates over the first guess

Answers to the Research Questions

Questions:

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

Fully coupled EnKF (with short windows) and 4D-Var (with longer 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 longer 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 longer 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.

Summary

- EnKF and 4D-Var are competitive, hybrid seems best
- EnKF being tested in many countries and labs.
- Ideas to further improve LETKF work well:
  - No-cost smoothing and "running in place" (K. and Yang, 2010, Penny)
  - Forecast sensitivity without adjoint model (Li and Kalnay, 2008)
  - Coarse resolution analysis without degradation (not shown) Yang et al
  - Model bias can be estimated and corrected (not shown) Danforth&K.
  - Adaptive inflation (helps a lot) can be combined with estimation of obs. errors (Miyoshi, 2011, Li et al. 2009).
  - Estimation of surface fluxes of carbon as evolving parameters seems to work well if several improvements are implemented (Kang et al)
- Coupled ocean-atmosphere analyses (Singleton, 2011)
  - 4D-Var and EnKF work well in a simple fully coupled ocean-atm. model. EnKF optimal for short windows, 4D-Var for long windows.
  - Optimal accuracy similar for both methods.
  - ECCO (4D-Var with fluxes as control variables, very long window) gives good analyses but not so good surface fluxes

New: Effective Assimilation of Precipitation (Guo-Yuan Lien, E. Kalnay and T Miyoshi)!

- Assimilation of precipitation has been done by changing the model moisture *q* in order to make it rain as observed.
- Very successful during the assimilation: e.g. the North American Regional Reanalysis.
- However the model forgets about the changes almost immediately after the assimilation stops!
- We believe it is because what assimilation of precipitation should do is modify efficiently potential vorticity (PV), the variable that the model **will remember**.
- EnKF, in principle, 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, about 5 years ago, we tried assimilating precipitation observations in a SPEEDY model simulation and the results were very disappointing.
- Another problem is that EnKF assumes model and obs perturbations are Gaussian. And precipitation is NOT Gaussian!

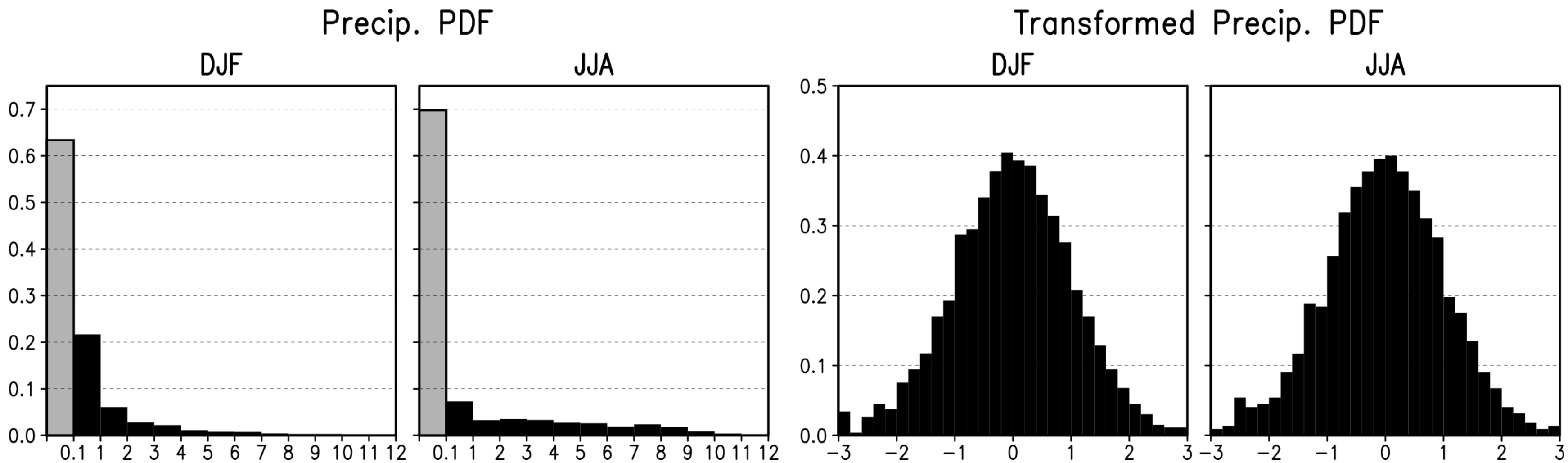


Figure 1:Figure 1: Examples of a 10-year climatology of precipitation and its Gaussian transform.

PV, however, to obtain a good EnKF assimilation of rain in our OSSEs. In addition we transform the observed and model precipitation variables into Gaussian distributions based on their climatology. This second problem is addressed by computing first the cumulative distribution function (CDF) $F(y)$ of a 10-years climatology of precipitation y for each grid point or observation station (Lien et al., 2012): $y_{trans} = G^{-1}[F(y)]$, where is the inverse CDF of the normal distribution. Figure 1 shows an example of the original precipitation and its Gaussian transform for summer and winter at a point in Maryland.

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

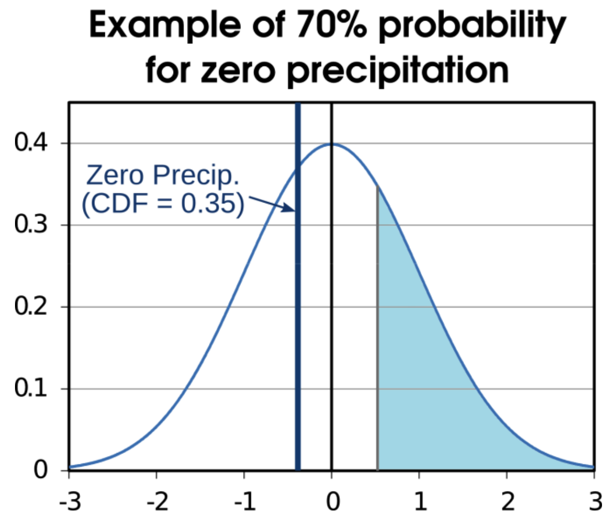


Figure 2

Figure 2: Example of the pdf value assigned to rain. If the probability of no rain is 70%, a no rain observation is assigned a CDF value of 0.35.

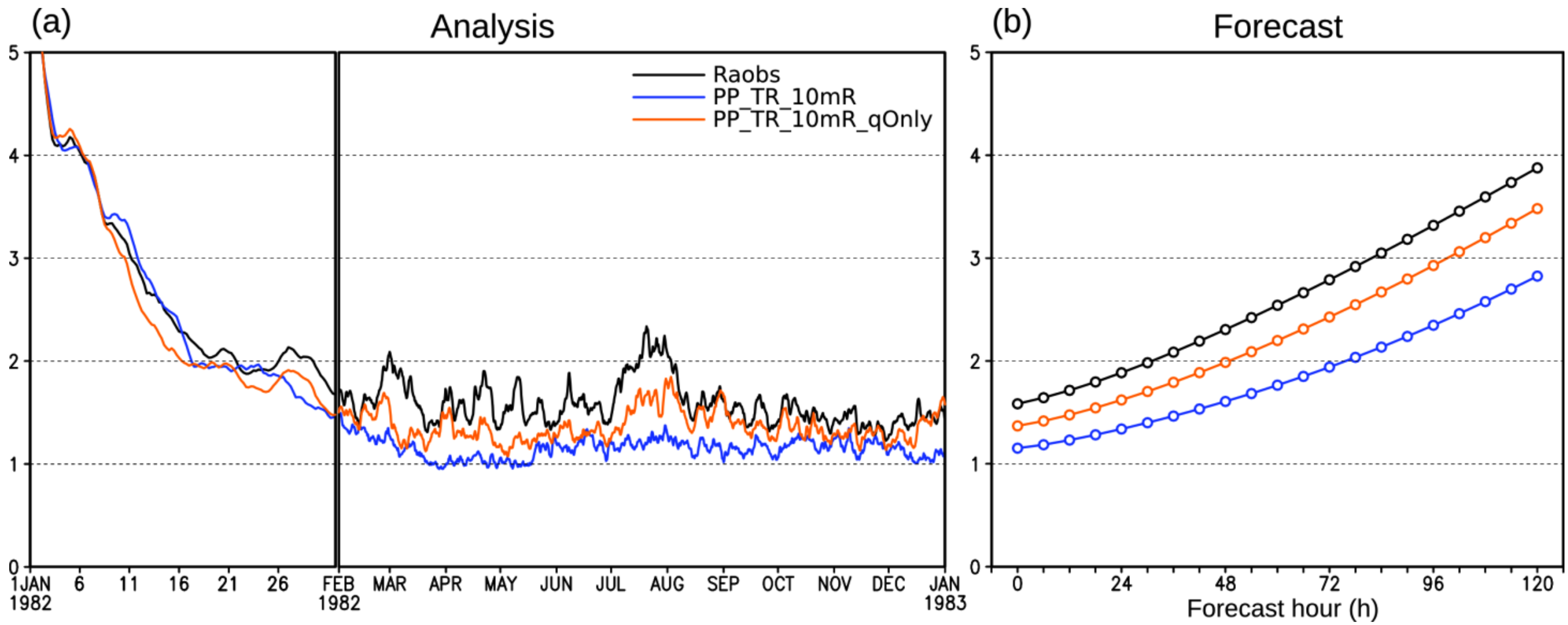

(a)
Analysis
Raobs
PP_TR_10mR
PP_TR_10mR_qOnly
1JAN 1982
FEB 1982
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
JAN 1983
(b)
Forecast
Forecast hour (h)

Figure 4

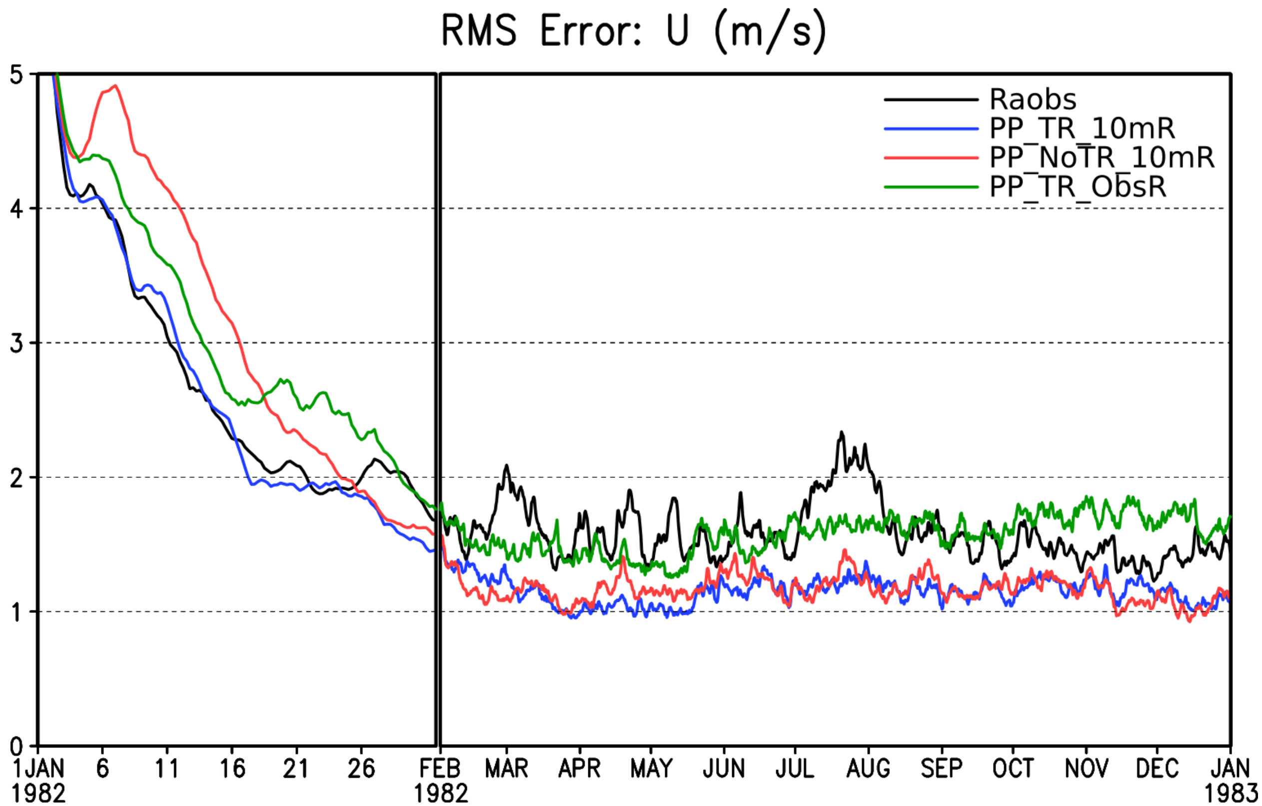

RMS Error: U (m/s)
Raobs
PP_TR_10mR
PP_NoTR_10mR
PP_TR_ObsR
5
4
3
2
1
0
1JAN 1982
6
11
16
21
26
FEB 1982
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
JAN 1983

Figure 5

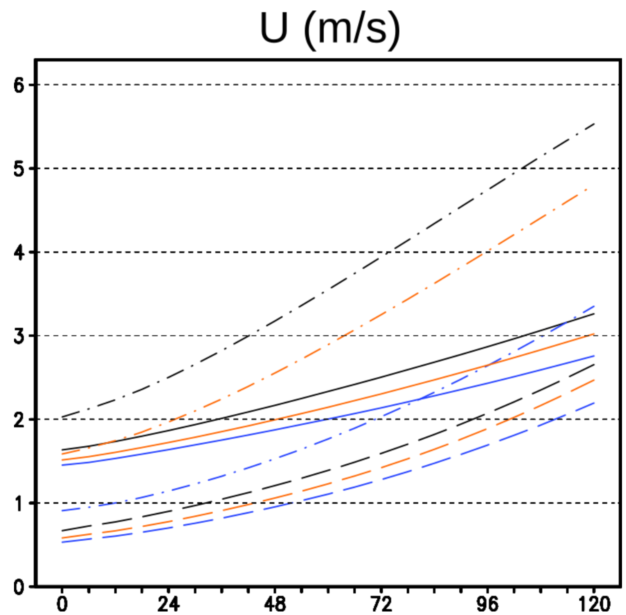


Figure 6

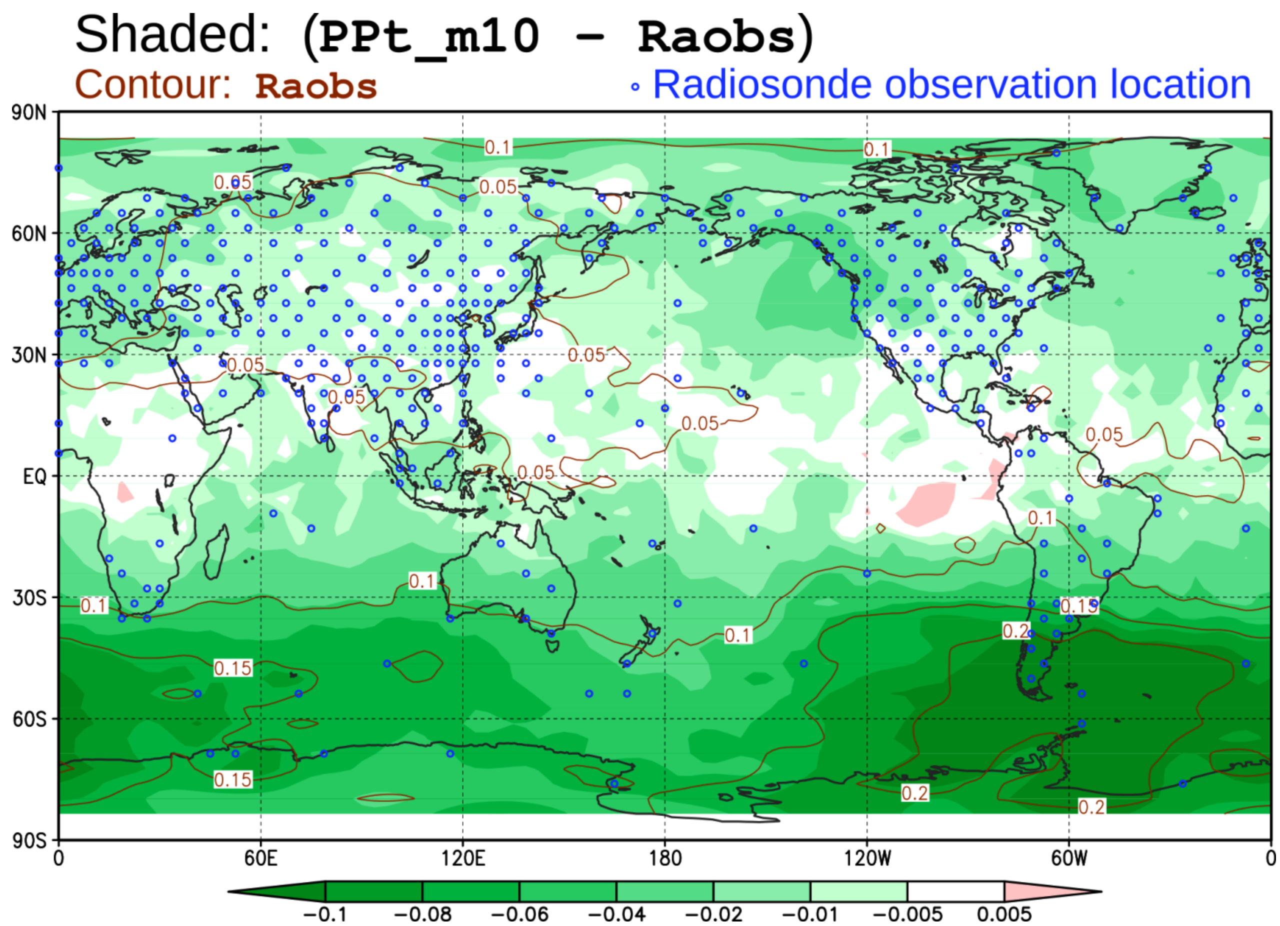
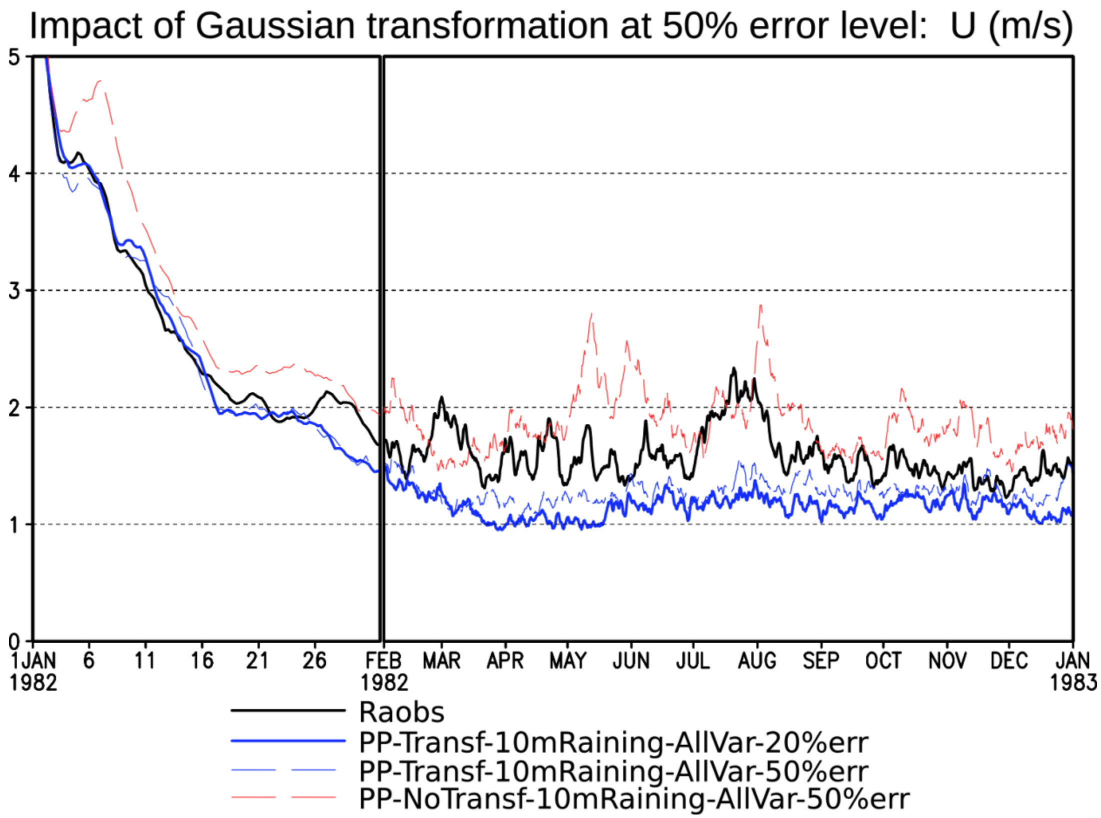

Shaded: (PPt_m10 - Raobs)
Contour: Raobs
Radiosonde observation location
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
−0.1
−0.08
−0.06
−0.04
−0.02
−0.01
−0.005
0.005

Figure 7



Impact of Gaussian transformation at 50% error level: U (m/s)
5
4
3
2
1
0
1JAN 1982
6
11
16
21
26
FEB 1982
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
JAN 1983
Raobs
PP-Transf-10mRaining-AllVar-20%err
PP-Transf-10mRaining-AllVar-50%err
PP-NoTransf-10mRaining-AllVar-50%err

Figure 8

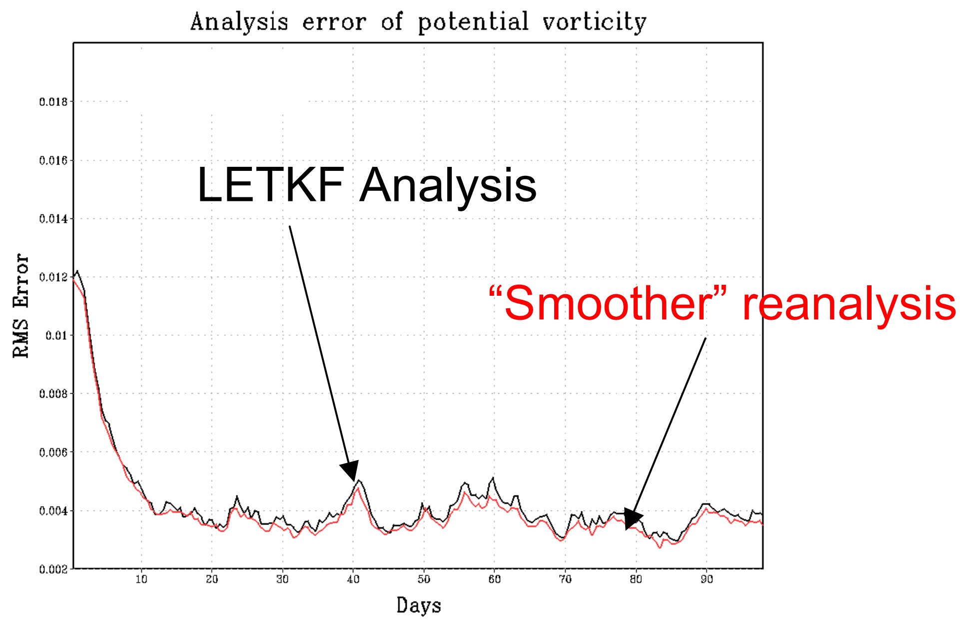
Add-on: no-cost LETKF smoother allows a comparison of EnKF initial and final increments: the initial 4D-Var increments are sensitive to the norm, the final increments are similar to EnKF

LETKF analysis
at time n

$$\overline{\mathbf{x}}_n^a = \overline{\mathbf{x}}_n^f + \mathbf{X}_n^f \overline{\mathbf{w}}_n^a$$

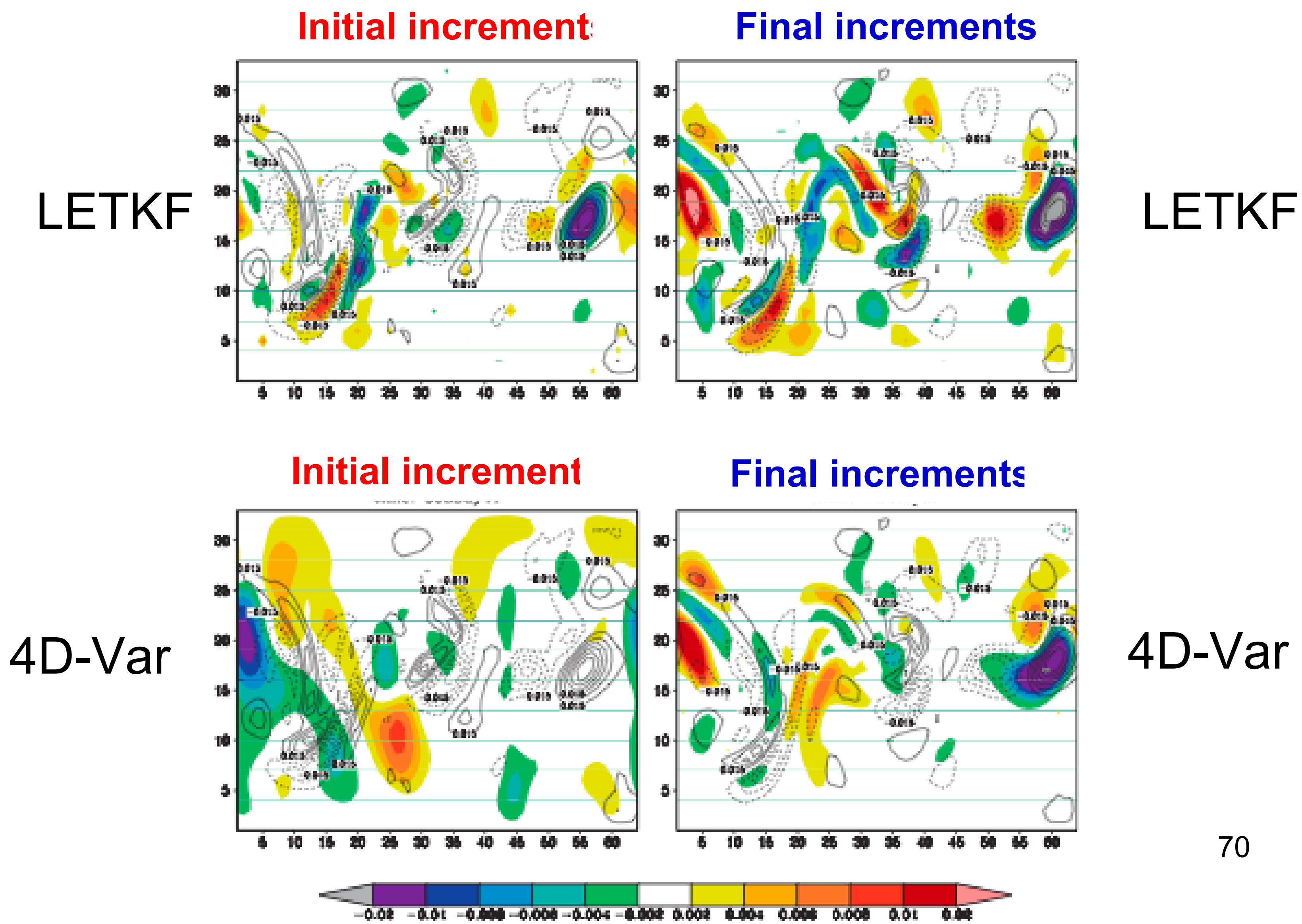
Smoother analysis
at time n-1

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



This very simple smoother allows us to go back and forth in time within an assimilation window: it allows assimilation of **future** data in reanalysis

Initial and final analysis corrections (colors), with one Bred Vector (contours)



Transformation method

- The "Gaussian anamorphosis" (similar to Schöniger et al. 2012 in hydrology):

  $$y_{trans} = G^{-1}[F(y)]$$

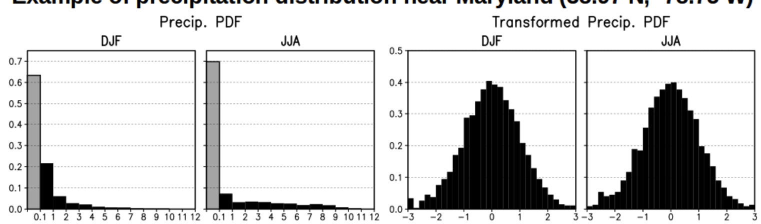
  y : precipitation variable

  F : Cumulative distribution function (CDF) of precipitation variables based on the 10-year model climatology at each grid and each season.

  G^{-1} : Inverse CDF of normal distribution.

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

Example of precipitation distribution near Maryland (38.97 N, -78.75 W)



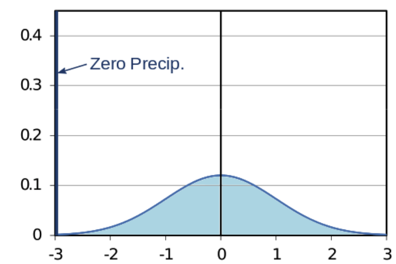
- LETKF is performed on the transformed space.
  - Variables that are needed to be transformed: $\mathbf{y}_{\mathrm{pp}}^{b(i)}$, $\bar{\mathbf{y}}_{\mathrm{pp}}^{b}$, $\mathbf{y}_{\mathrm{pp}}^{o}$.
  - The observation errors associated with each observation also have to be transformed. Conceptually:

  $$\sigma^o_{trans} \simeq (y^o+\sigma^o)_{trans} - y^o_{trans} \simeq y^o_{trans} - (y^o-\sigma^o)_{trans}$$

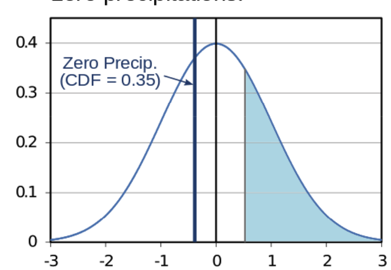
Handling the zero-precipitation data

Example of 70% probability for zero precipitation

(a) Consider only the non-zero precipitation value in the Gaussian transformation.



(b) Consider all precipitation data in the Gaussian transformation, assign ***middle values*** of no-rain cumulative probability for all zero precipitations.



(c) Consider all precipitation data in the Gaussian transformation, assign ***uniformly distributed random values*** within no-rain cumulative probability for all zero precipitations.