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), and

Eugenia Kalnay

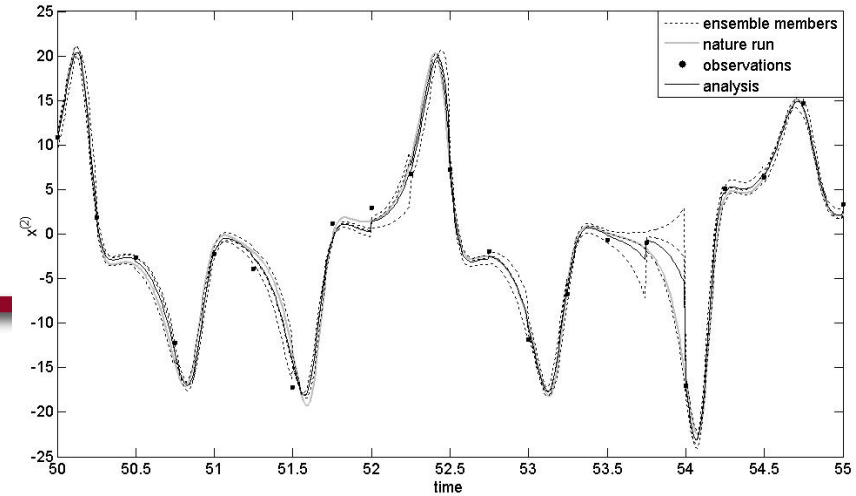
University of Maryland

UMD Weather-Chaos Group: **Kayo Ide**, **Brian Hunt**, Ed Ott, and students (Javier Amezcua, Yan Zhou, Ying Zhang)

Also: Malaquías Peña, Matteo Corazza.

JCSDA, 13 September 2011

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. Requires model adjoint. :-(
  - ✓ Ensemble Kalman Filter, 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, Japan, NCEP, ECMWF, Canada, Brazil, Germany… are now exploring EnKF (or hybrid EnKF+variational) for operations.

This talk: tools that improve EnKF

We adapted ideas inspired by 4D-Var:

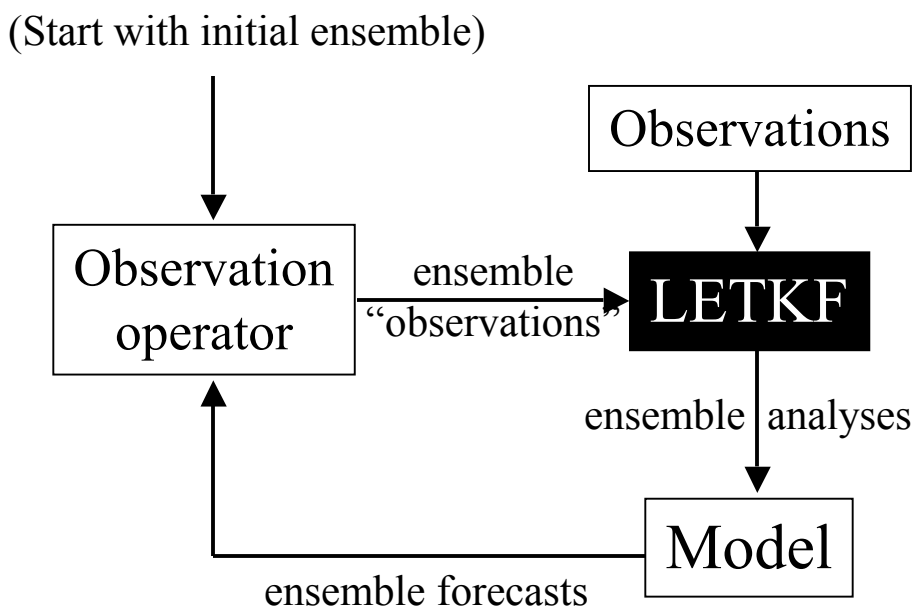
- ✓ **No-cost smoother** (Kalnay et al, Tellus 2007)
- ✓ "**Running in Place**" and "**Quasi Outer Loop**", deal with **spin-up**, nonlinearities and long windows (Kalnay and Yang, QJ 2010, Yang and Kalnay, MWR 2011)
- ✓ **Forecast sensitivity** to observations (Liu and Kalnay, QJ, 2008) (correction)
- ✓ **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:

- ✓ Estimates of surface carbon fluxes as parameters (Kang et al, 2011)
- ✓ Global Ocean Data Assimilation (Penny, PhD thesis, 2011)

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

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

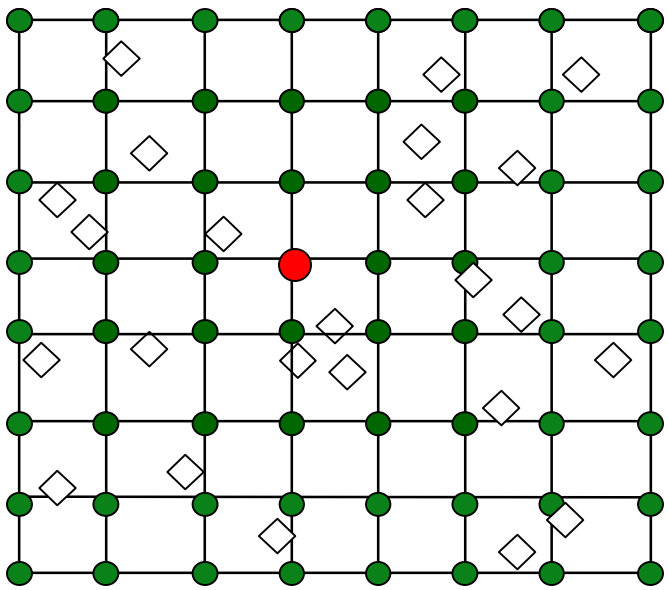


- Model independent (black box)
- Obs. assimilated simultaneously at each grid point
- 100% parallel
- No **adjoint** needed
- 4D LETKF extension
- Computes **weights** explicitly

Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

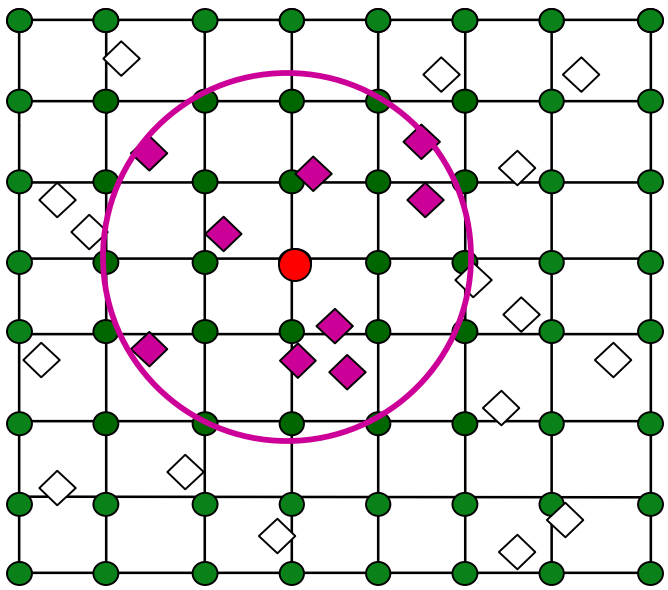


Localization based on observations

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 $\mathbf{X}^b = \left[\mathbf{x}^b_1 - \bar{\mathbf{x}}^b \mid \ldots \mid \mathbf{x}^b_K - \bar{\mathbf{x}}^b\right];$

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

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

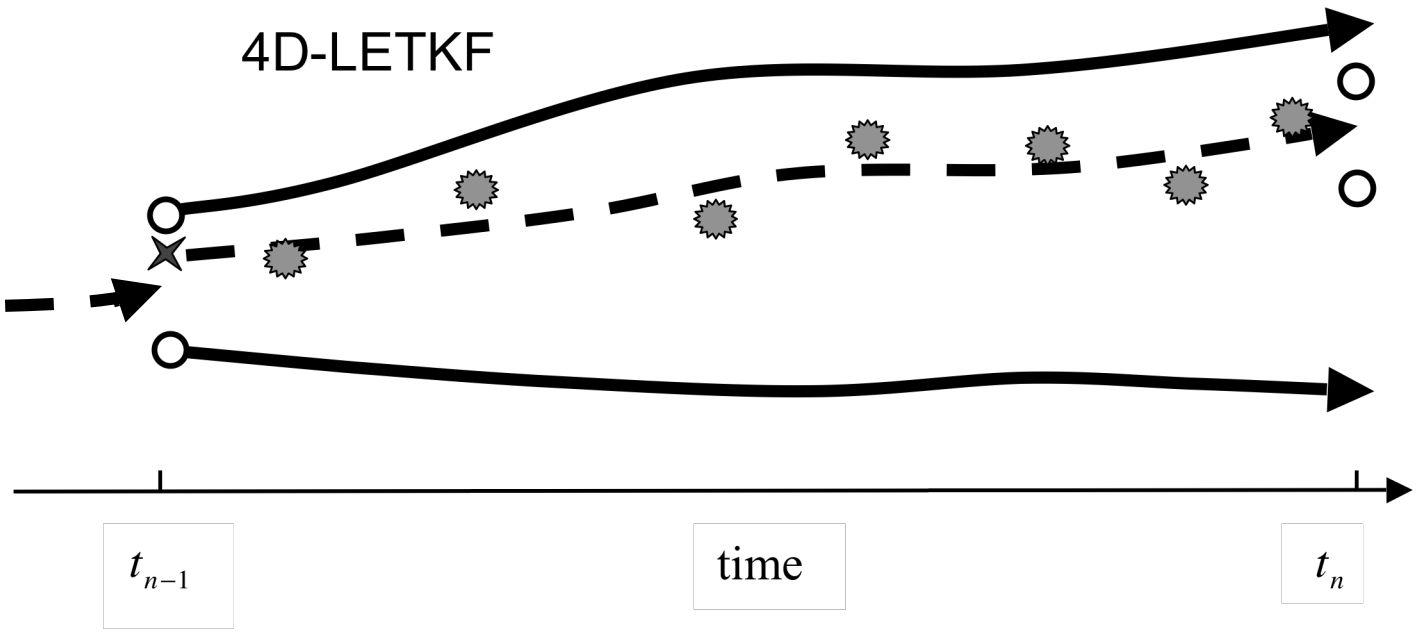
$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1};\ \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

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

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 $\mathbf{X}^a_n = \mathbf{X}^b_n\mathbf{W}^a + \bar{\mathbf{x}}^b$. Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights $\bar{\mathbf{w}}^a$ 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

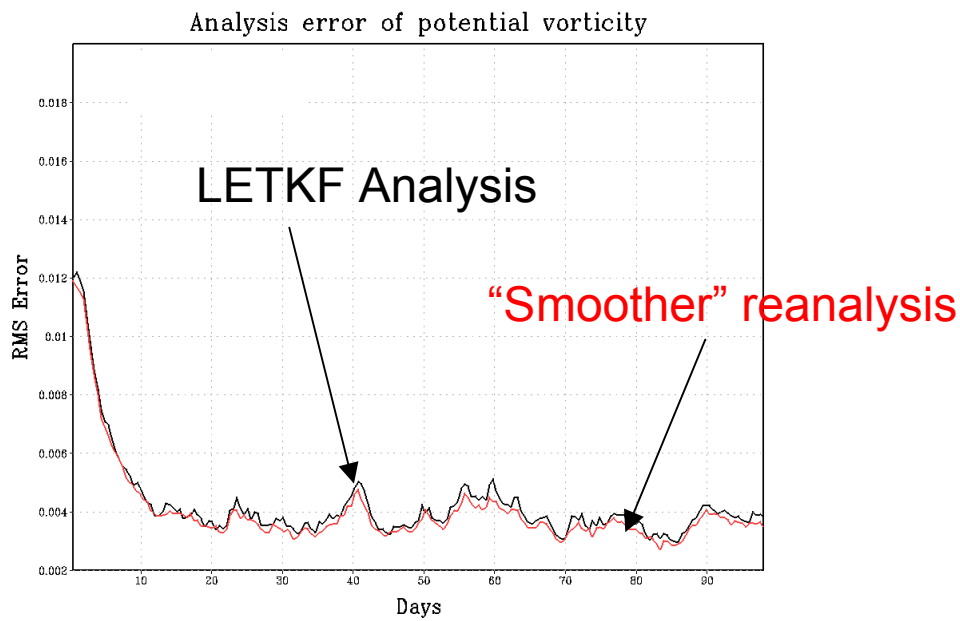
No-cost LETKF smoother tested on a QG model: it works…

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

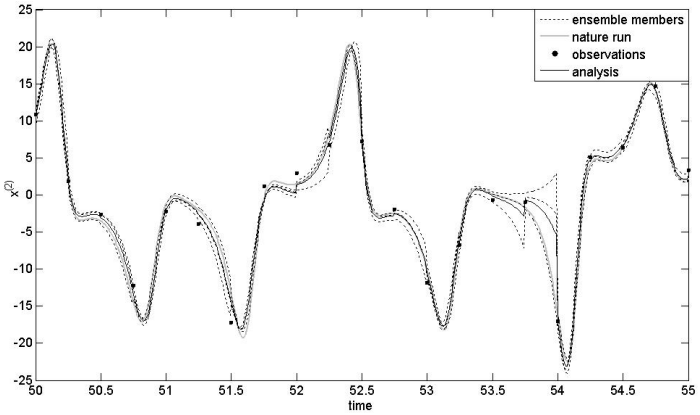
$$\tilde{\mathbf{x}}_{n-1}^a = \overline{\mathbf{x}}_{n-1}^f + \mathbf{X}_{n-1}^f \overline{\mathbf{w}}_n^a$$



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)

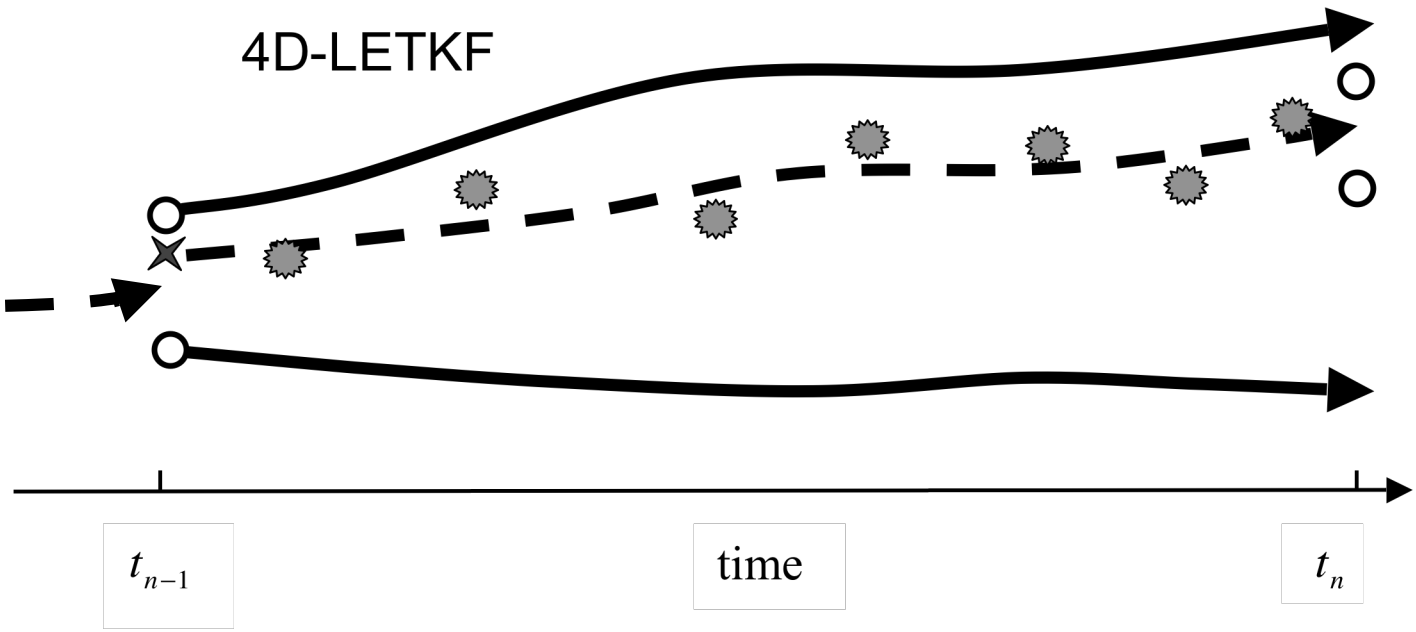


Lorenz -3 variable model
(Kalnay et al. 2007a Tellus),
RMS analysis error:

	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"

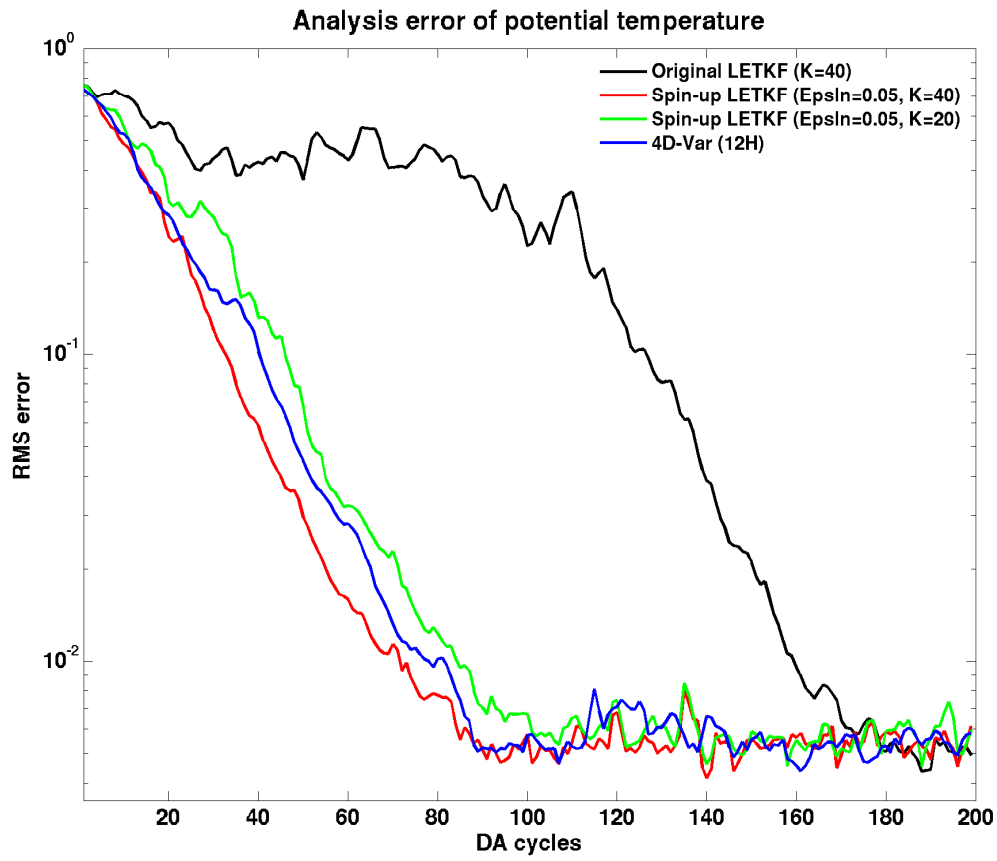
Quasi Outer Loop: similar to 4D-Var: use the final weights to correct only the <u>mean</u> initial analysis, keeping the initial perturbations. Repeat the analysis once or twice. It centers the ensemble on a more accurate nonlinear solution.

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	0.66	**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)

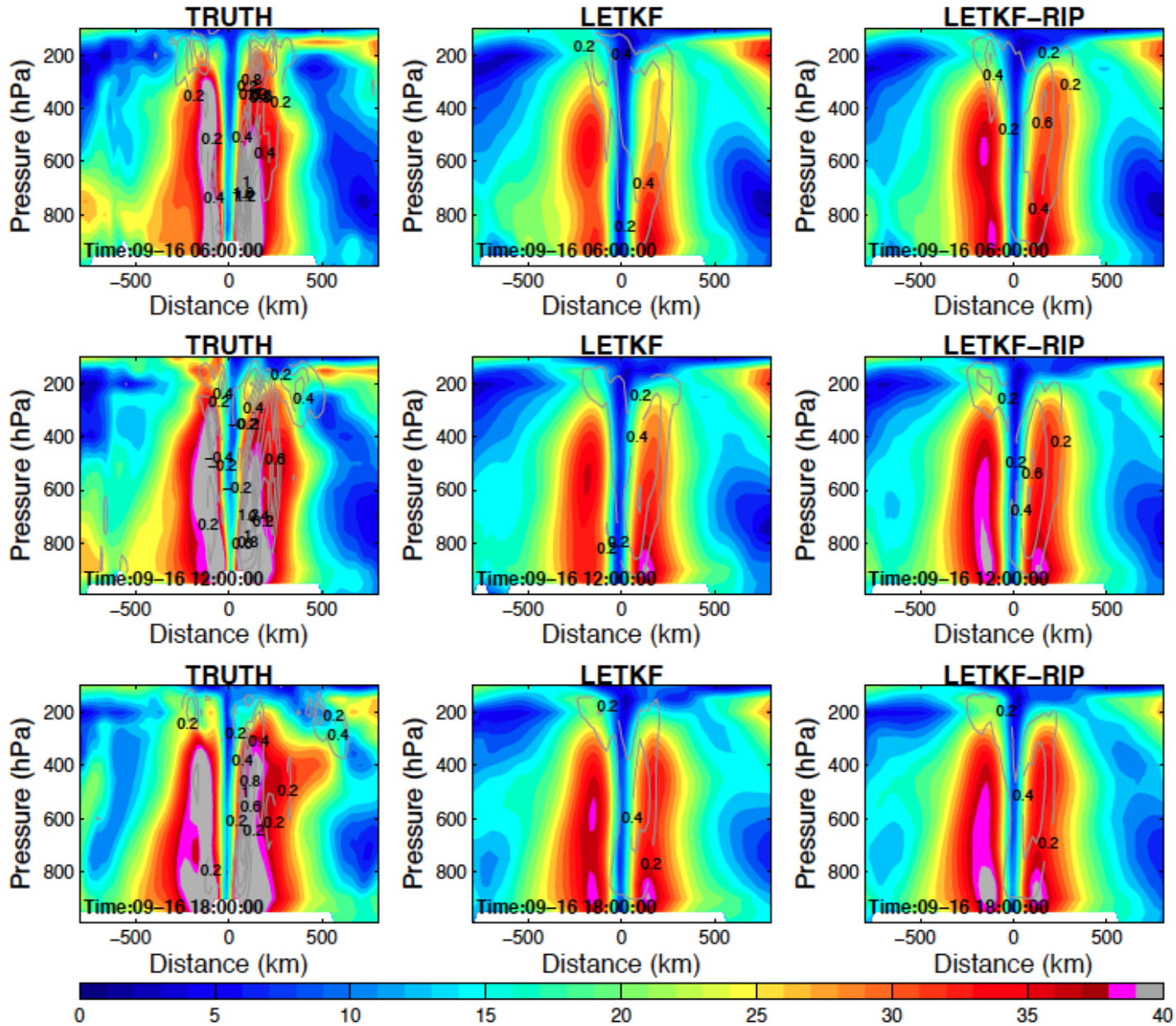
Spin-up depends on initial perturbations, but RIP works well even with random perturbations. It becomes as fast as 4D-Var (blue). RIP takes only 2-6 iterations.

Yang: 24 hr forecast of simulated Typhoon Sinlaku (trajectory and intensity were both improved with RIP)

T_{init}= 09/15 06Z

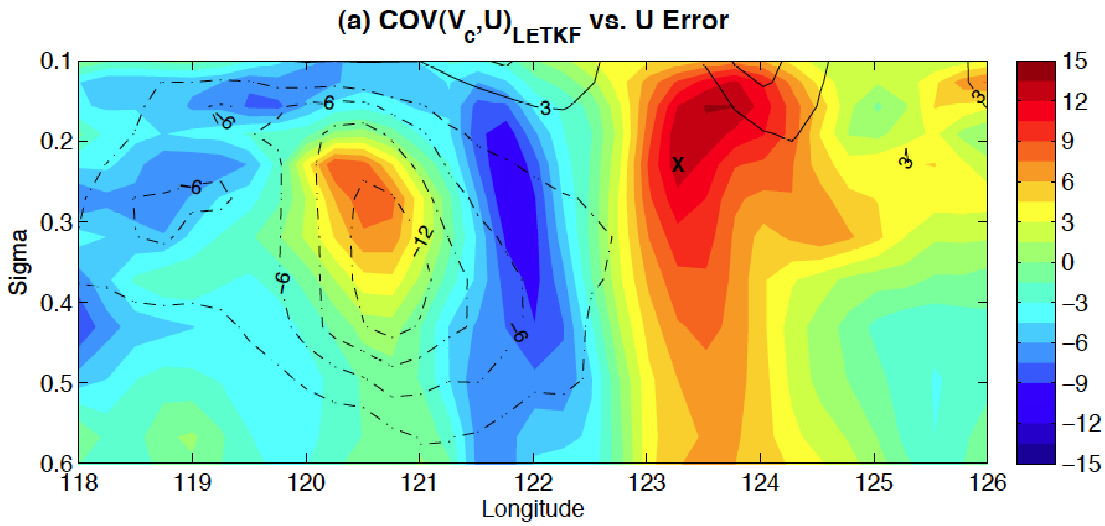
T_{init}= 09/15 12Z

T_{init}= 09/15 18Z

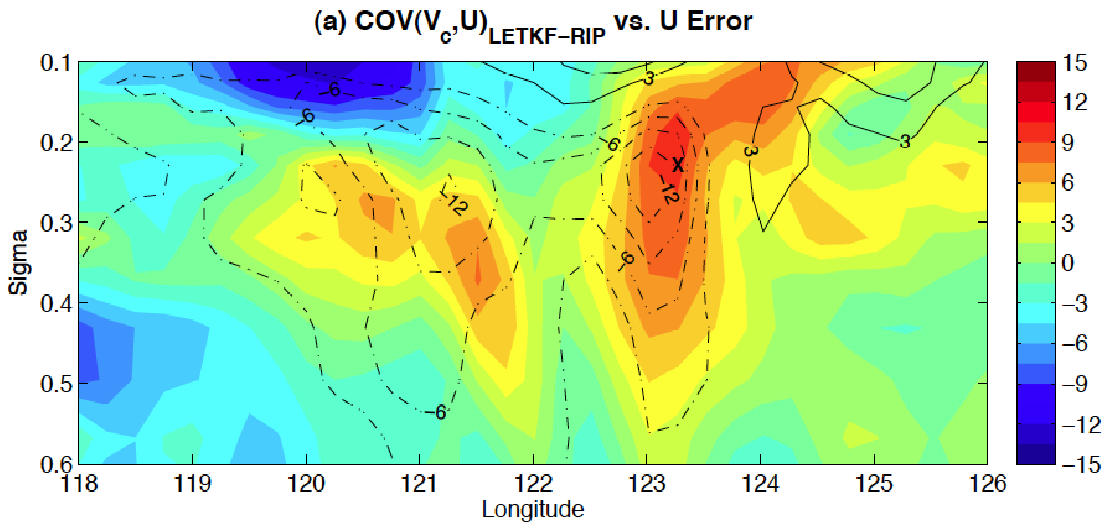


Typhoon Sinlaku: RIP improves the error covariance

(Yang et al., 2011)



With LETKF, the error covariance B (contours) is not related to the actual errors (colors) during spin-up.



With LETKF-RIP, the error covariance B represents better the the “errors of the day”. As a result the spin-up is faster and the forecast improves.

An application of LETKF-RIP to ocean data assimilation

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

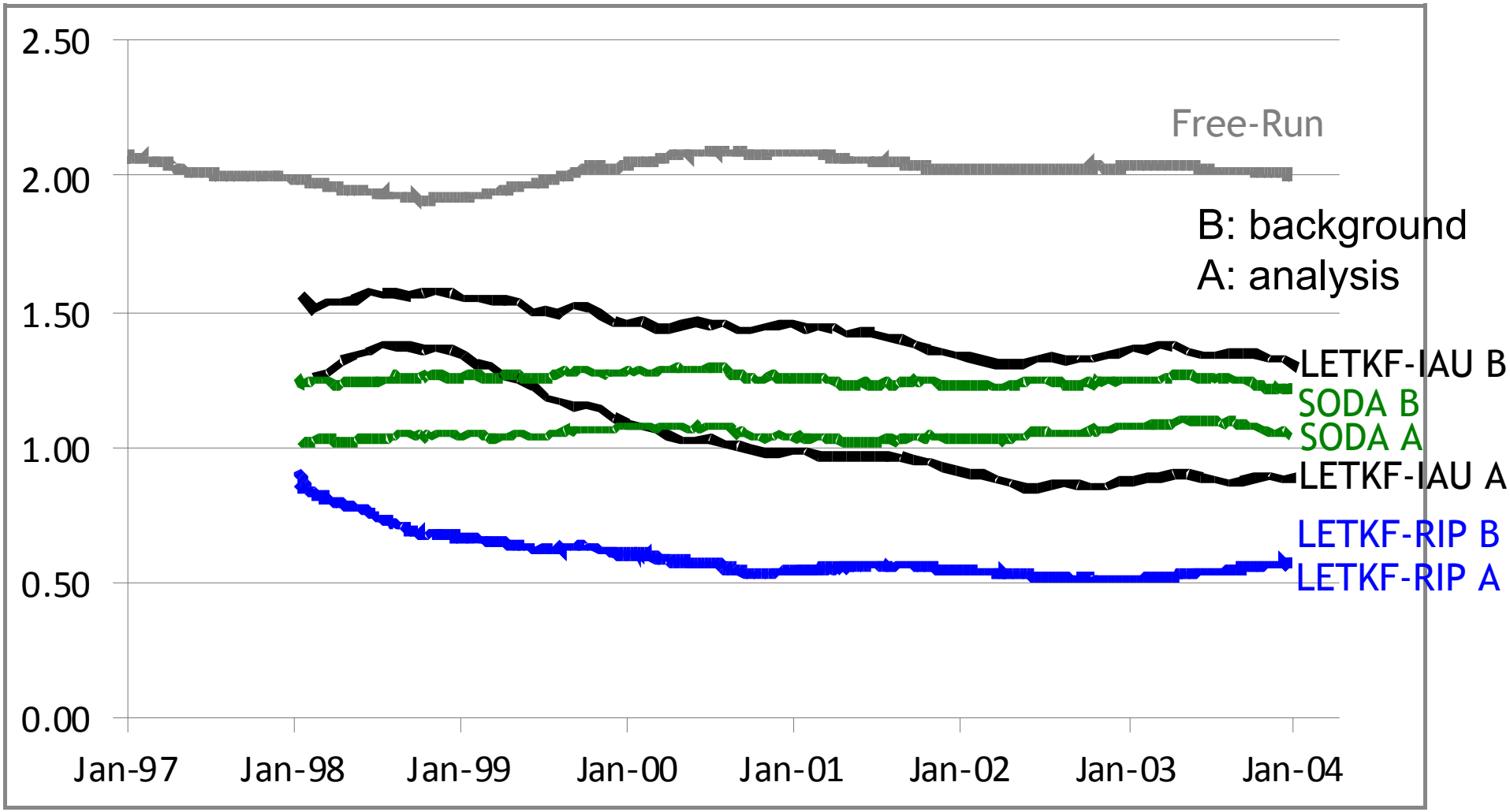
Steve Penny's thesis defense

April 15, 2011

Advisory Committee: 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)

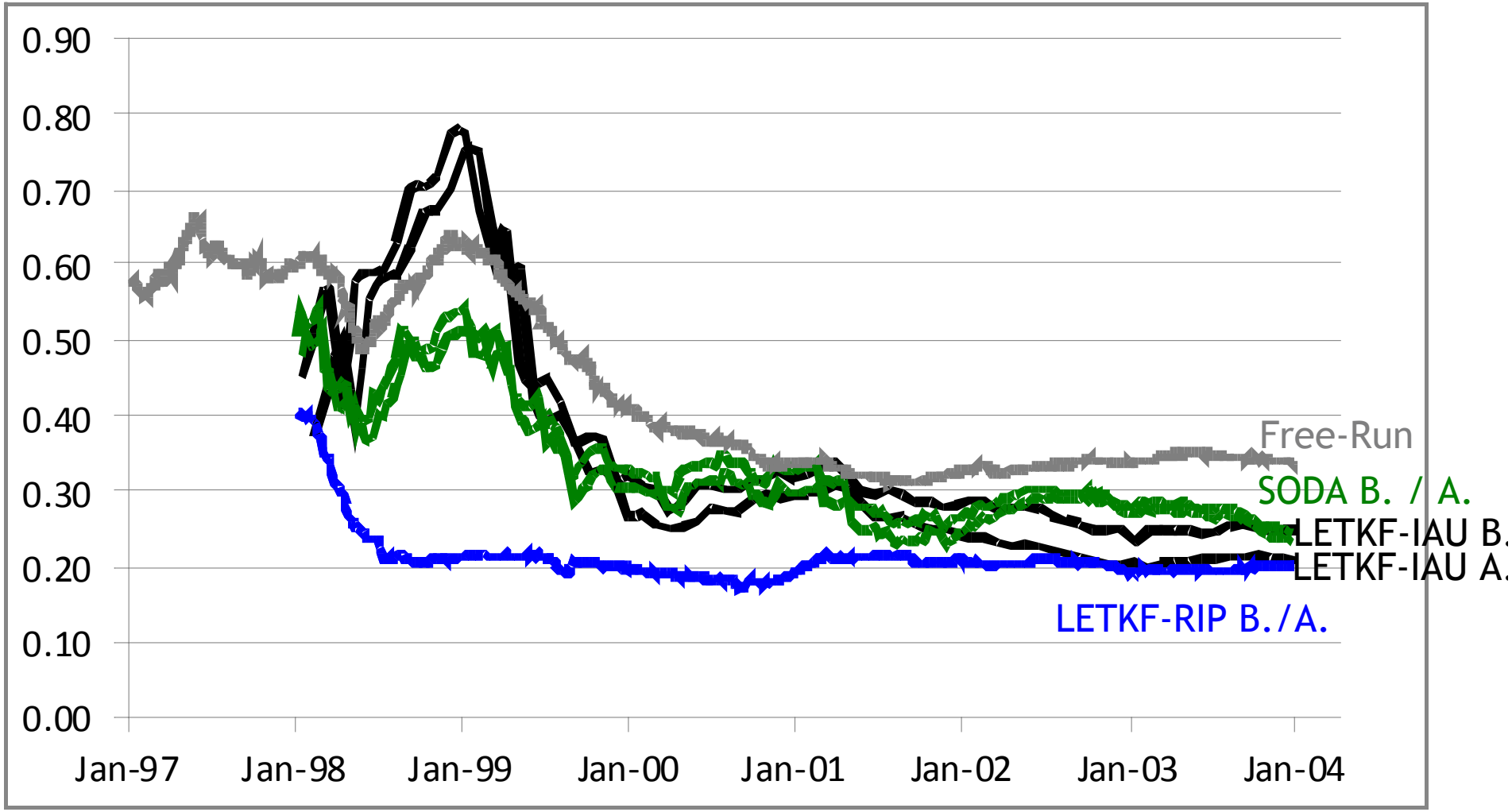
RMSD (ºC) (All vertical levels) 7 years of Reanalysis



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

RMSD (psu) (All vertical levels)

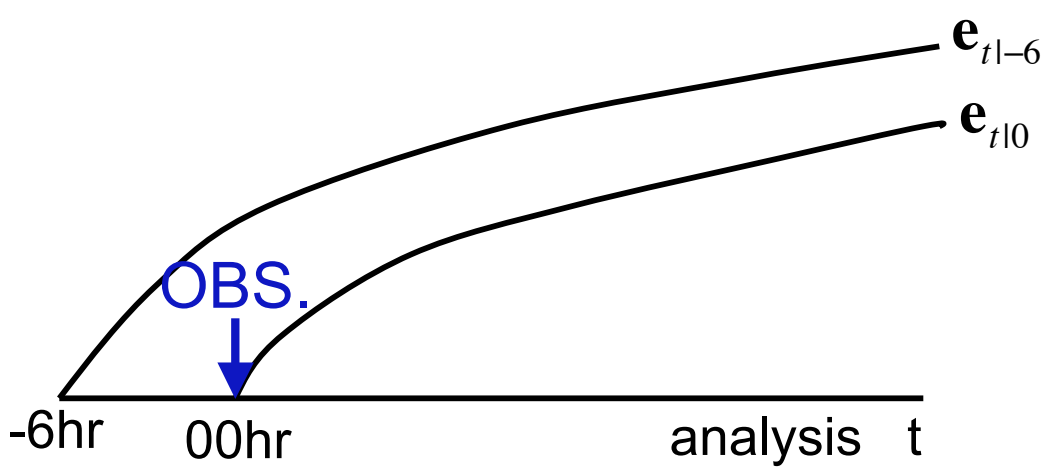
7 years of Reanalysis



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)



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

(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

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

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

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

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

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

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

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

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:

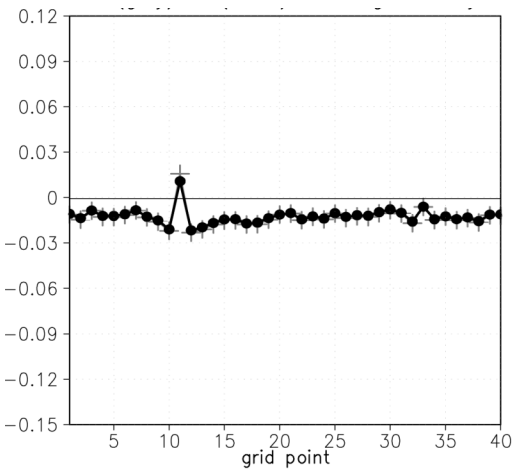
$$\Delta \mathbf{e}^2_P = \left[(\mathbf{y} - H(\mathbf{x}^b_{0|-6}) \right]^T \mathbf{R}^{-1} \mathbf{Y}^a_0 \mathbf{X}^{fT}_{t|0} \mathbf{P}^T \mathbf{P} (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) / (K-1)$$

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

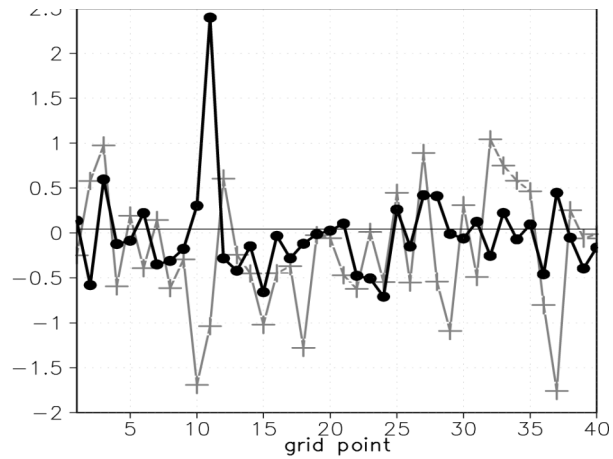
Observation impact from LB(+) and from ensemble sensitivity (•)

(Liu, pers. comm.)

1 day



10 days



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

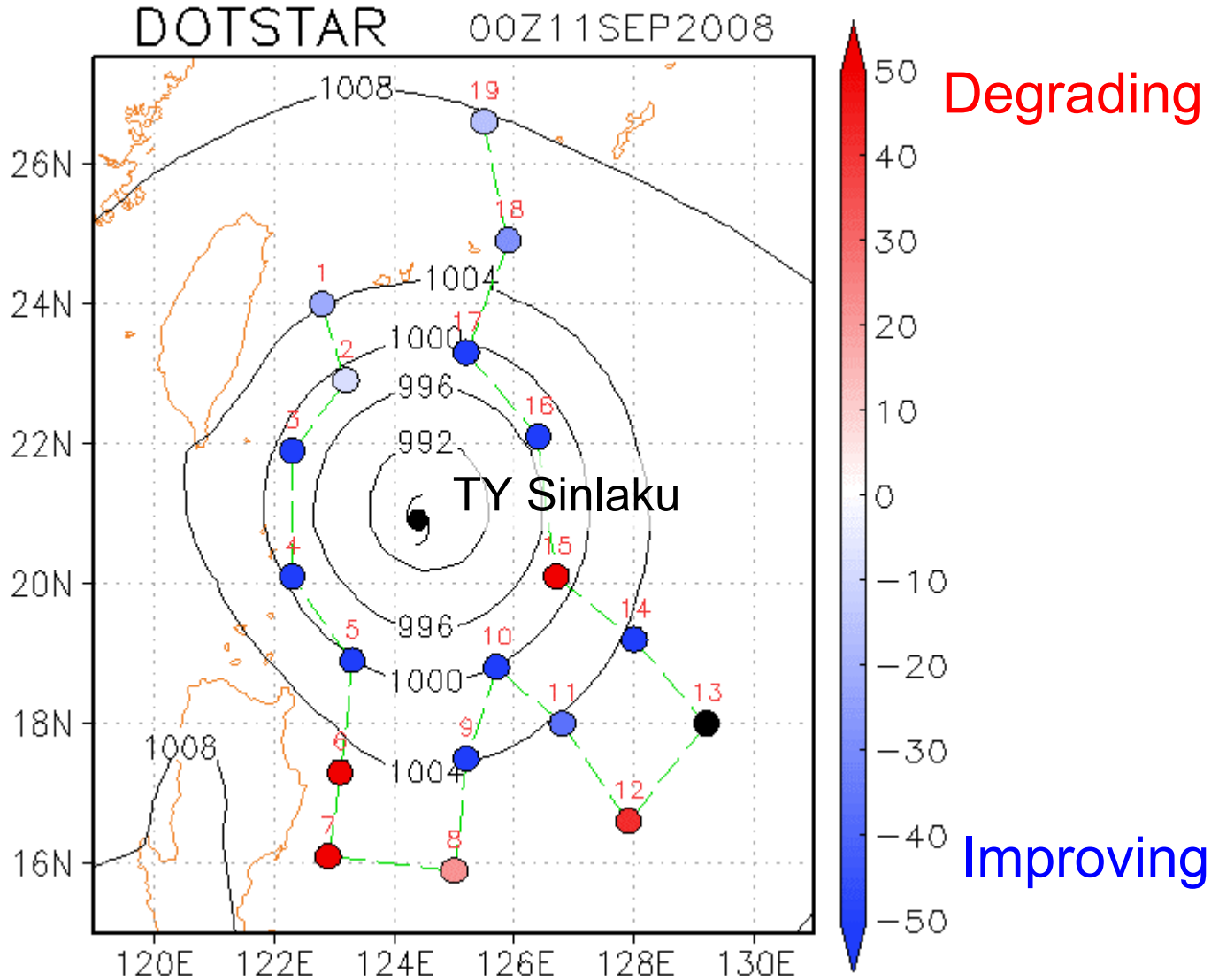
✓The ensemble sensitivity is nonlinear and is able to detect bad obs for longer forecasts.

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

Impact of dropsondes on a Typhoon

Estimated observation impact

Kunii et al., 2011

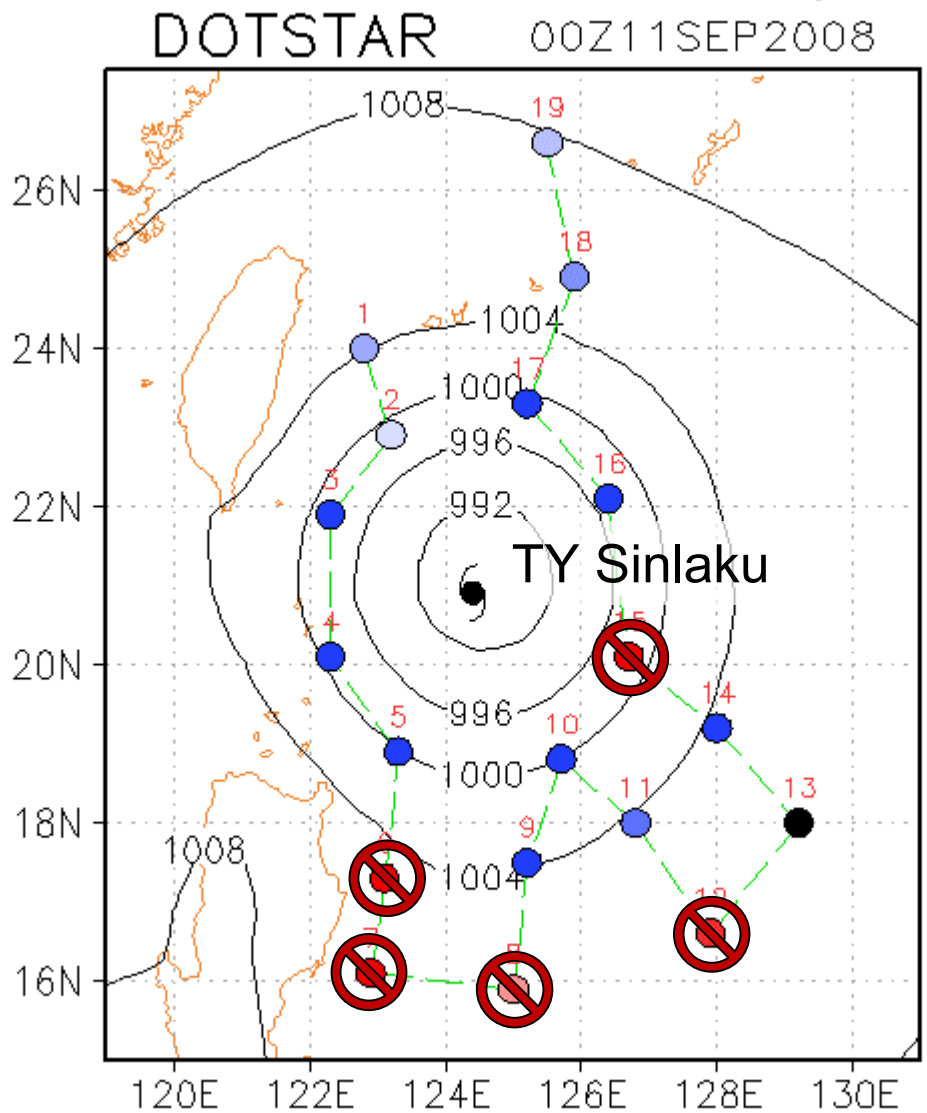


Degrading

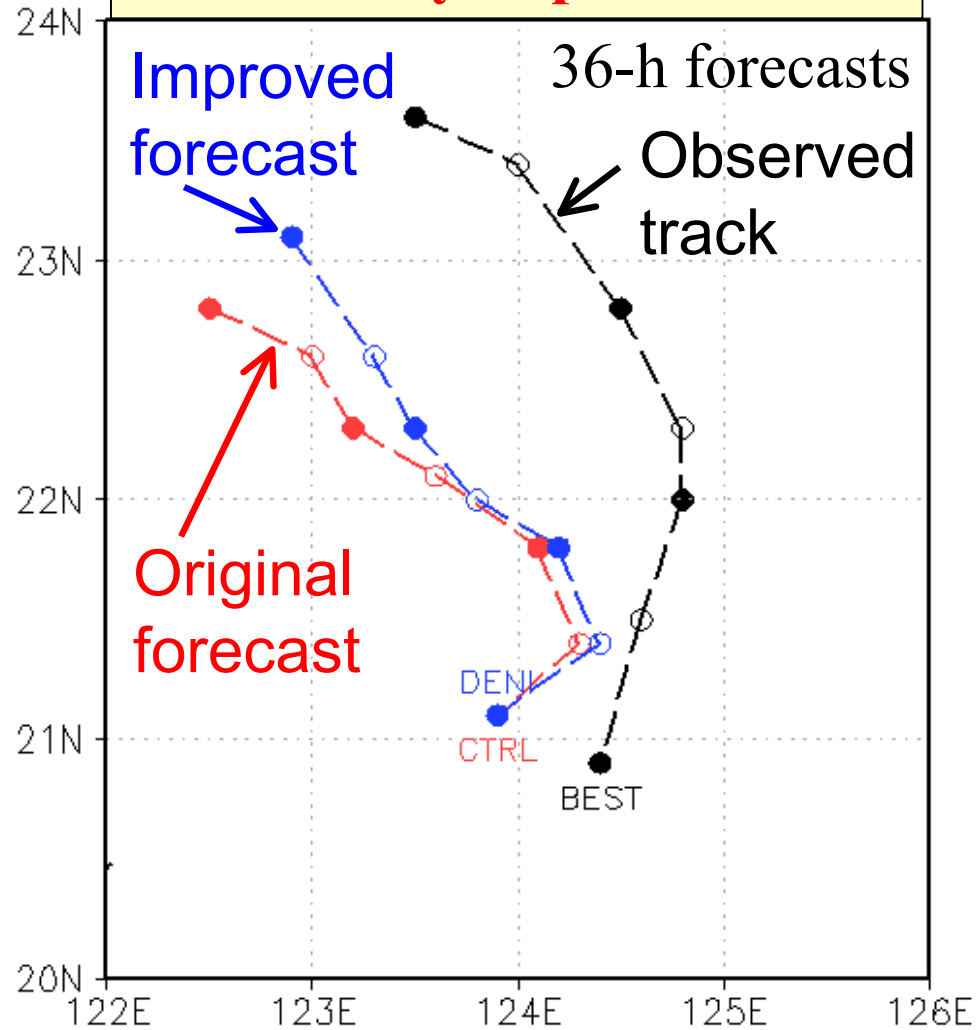
Improving

Denying negative impact data improves forecast!

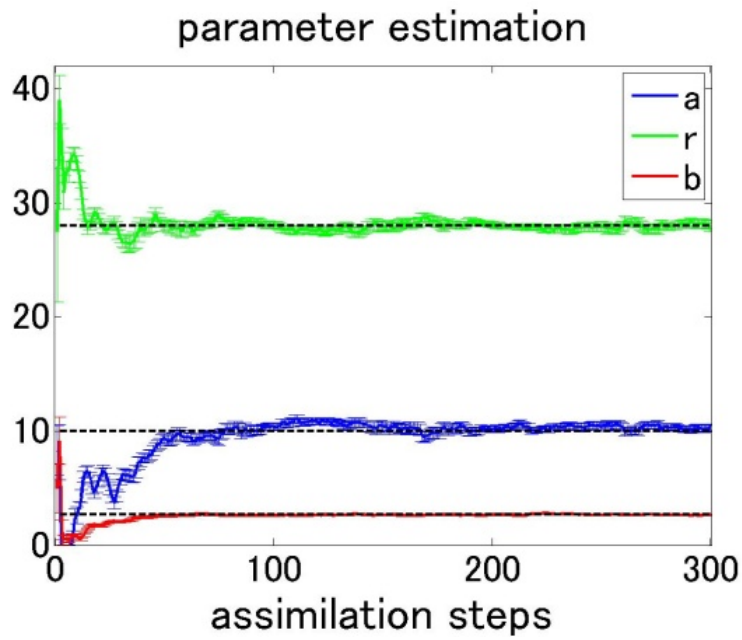
Estimated observation impact

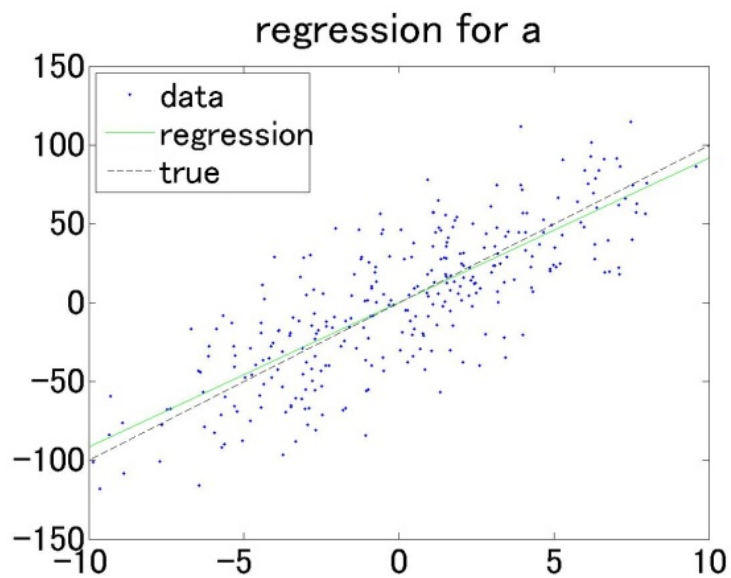


Typhoon track forecast is actually improved!!



Application: parameter estimation in EnKF. The state vector is augmented with the parameters *a*,*r*,*b*





Example with Lorenz model (simulation with noisy obs., courtesy: Miyoshi)

<u>Left:</u> estimation of parameters with LETKF from the error covariance; <u>Right</u>: standard regression from observations

Application: Parameter Estimation in EnKF

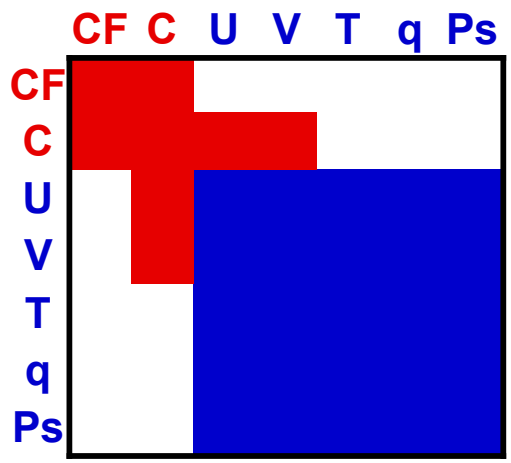
- Example of **carbon cycle data assimilation**
  - Surface CO_2 fluxes (CF): the forcing for atmospheric CO_2
- State vector augmentation
  - State vector is augmented by CF which is **updated by error covariance between the variables in the state vector**
- Variable localization
  - In a multivariate analysis of EnKF, **error covariance is zeroed out when there is no significant physical relationship between variables, in order to reduce a sampling error**
- Adaptive Multiplicative Inflation and Additive Inflation
  - They help represent background uncertainty more accurately
- Vertical localization of satellite column data
  - Averaging kernel is nearly uniform in the vertical, although the forcing term (our ultimate estimate) is at the surface.

Kang et al. (2011, JGR)

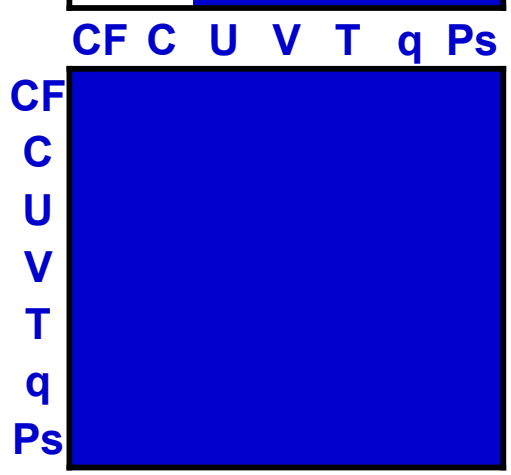
Variable localization

- Analysis of surface CO_2 fluxes assimilating atmospheric CO_2 observations
  - A case with a constant forcing

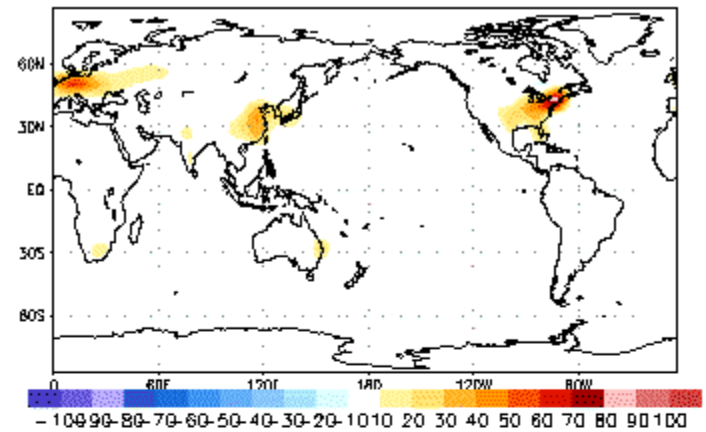
1-way multivariate analysis with variable localization ➔



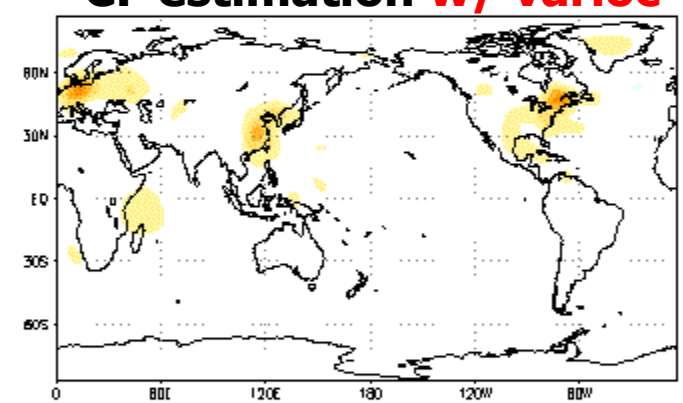
Fully multivariate analysis ➔



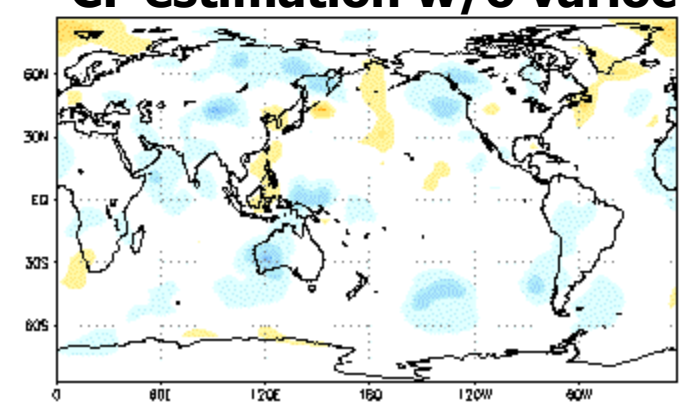
True fluxes



CF estimation w/ varloc



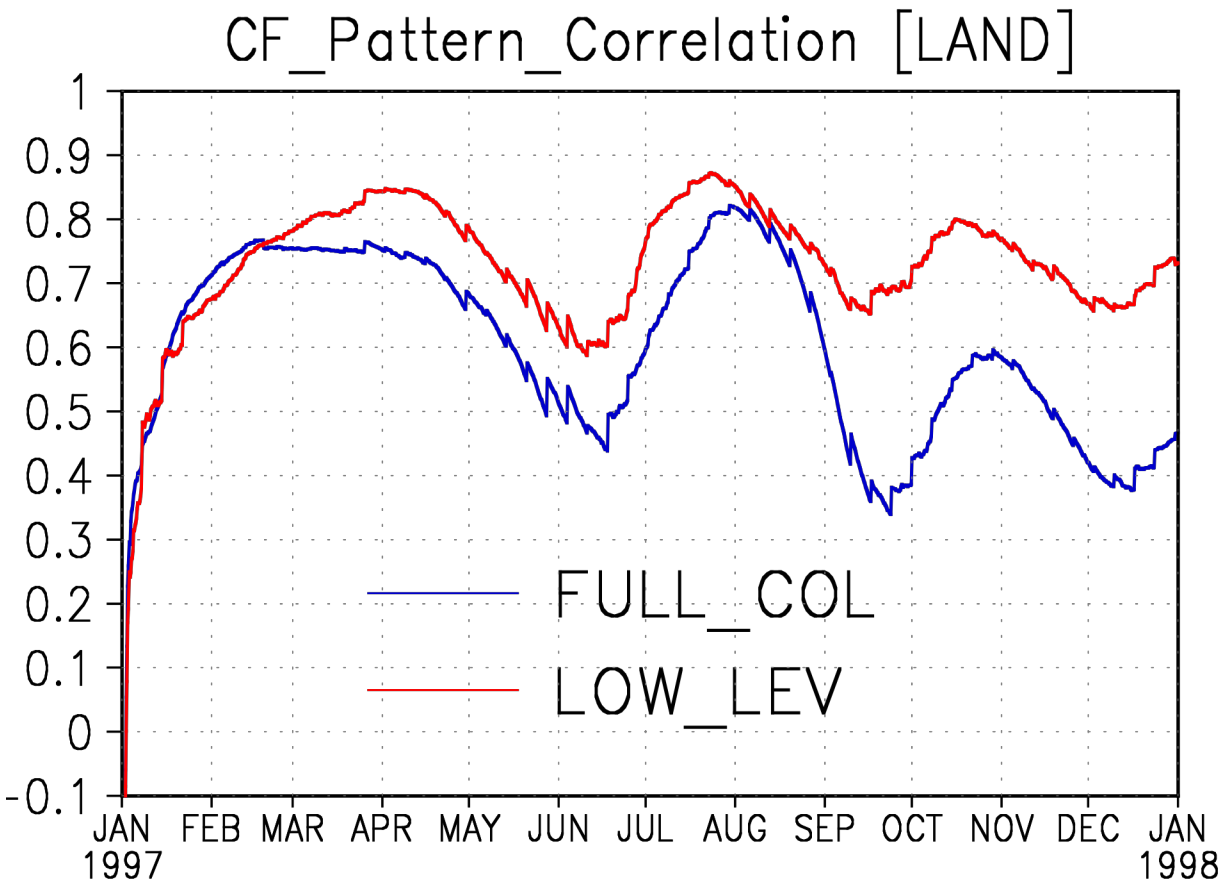
CF estimation w/o varloc

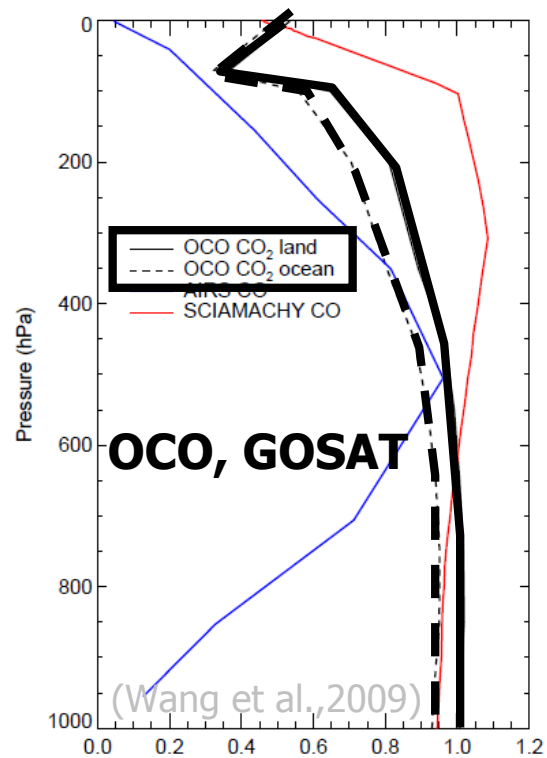


Kang et al. (2011, JGR)

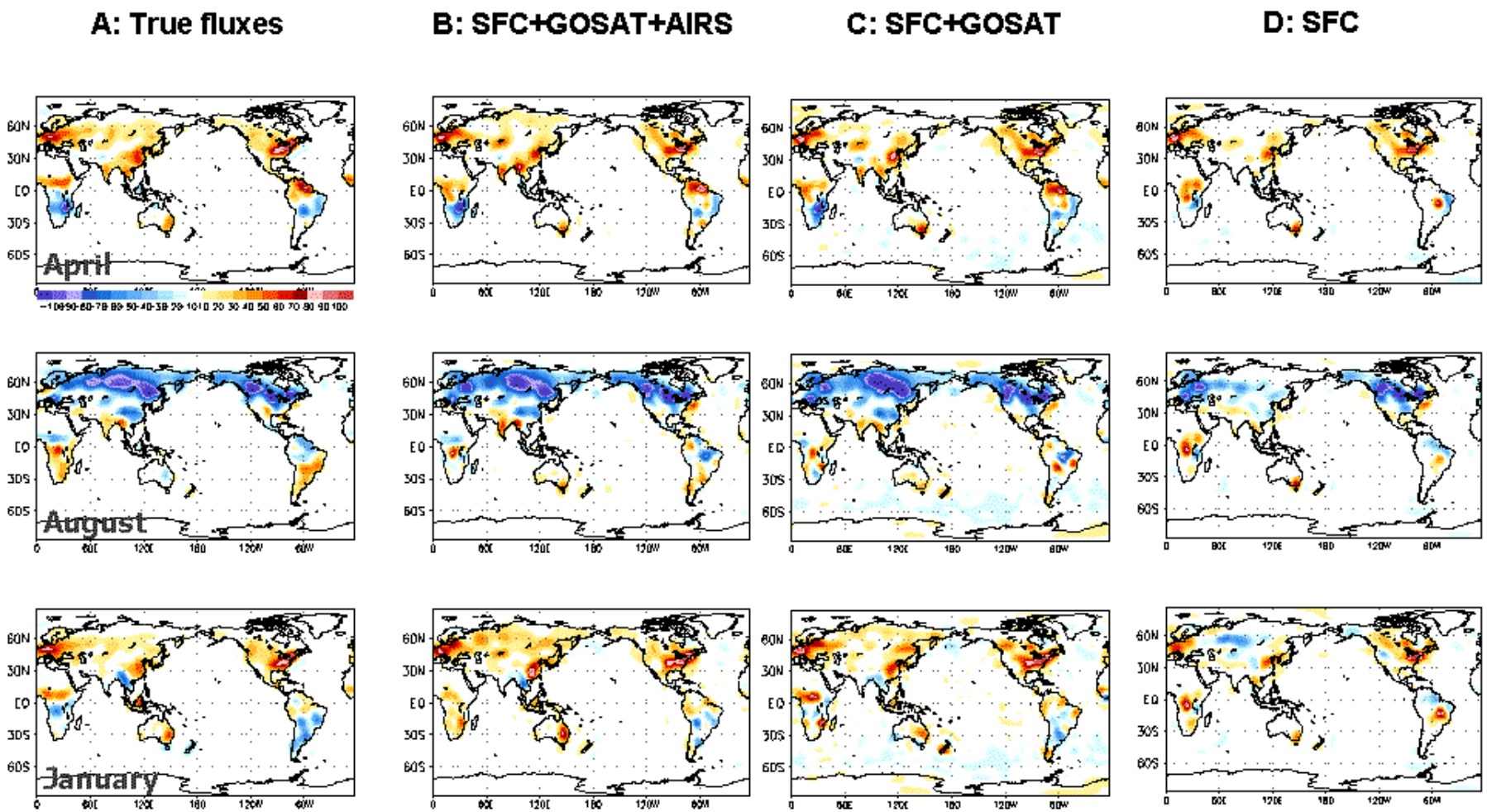
Vertical localization also improves results

Pattern correlation between CF estimate and true state





Surface carbon fluxes estimated with different CO2 observations



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

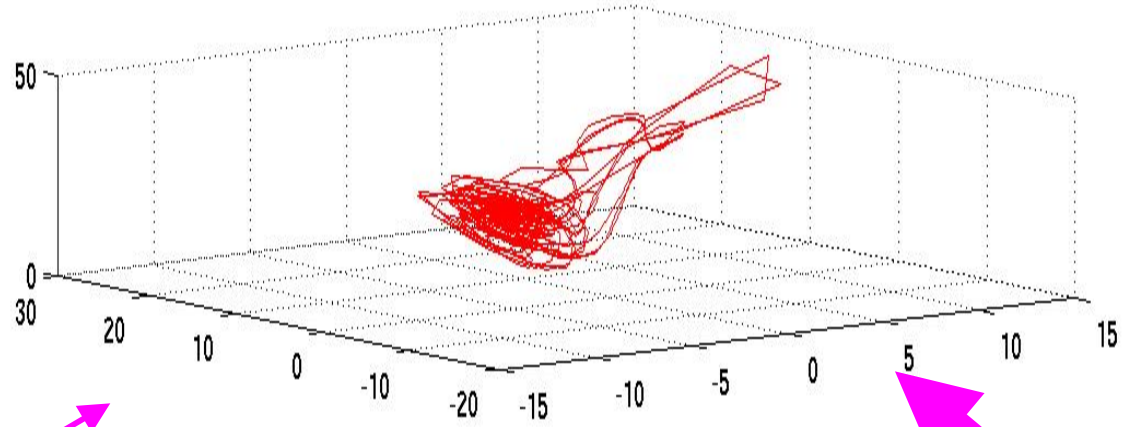
$$\dot{X} = \tau\sigma(Y - X) - c(x_t + k_2)$$
$$\dot{Y} = \tau rX - \tau Y - \tau XZ + c(y_t + k_2)$$
$$\dot{Z} = \tau XY - \tau bZ + c_z z_t$$

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)

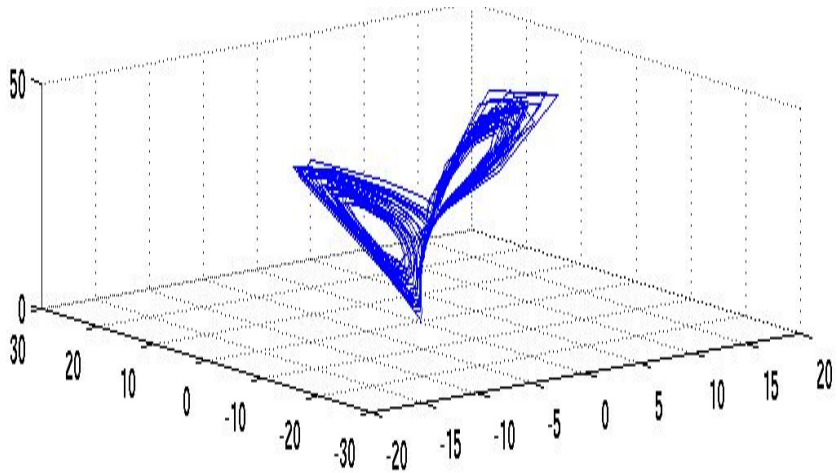
Tropical Atmosphere

Coupling strength

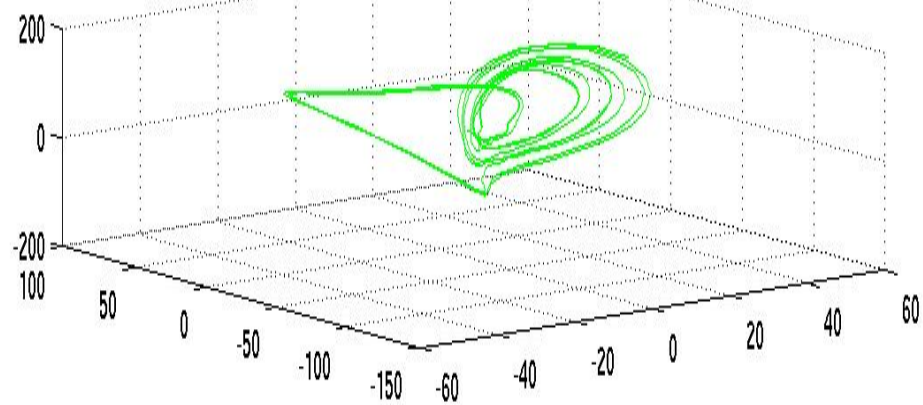


Ocean is vacillating between a "normal" (lasts about 3-12 years) and "El Nino" state (lasts about a 1 year)

Extra-tropical Atmosphere



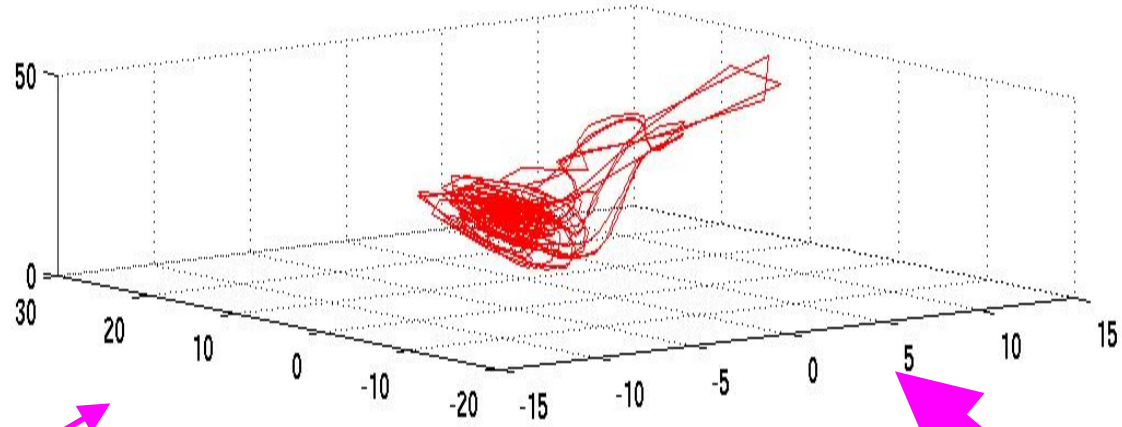
Tropical Ocean



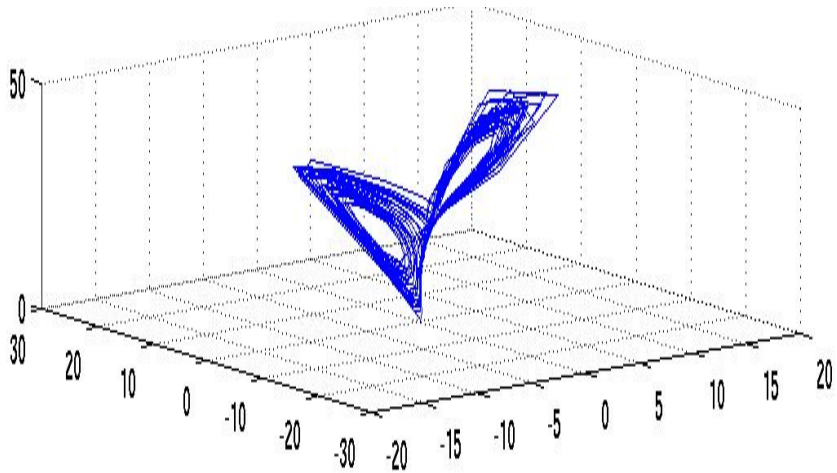
Simple Coupled Ocean-Atmosphere Model (Peña and Kalnay, 2004)

Tropical Atmosphere

Coupling strength

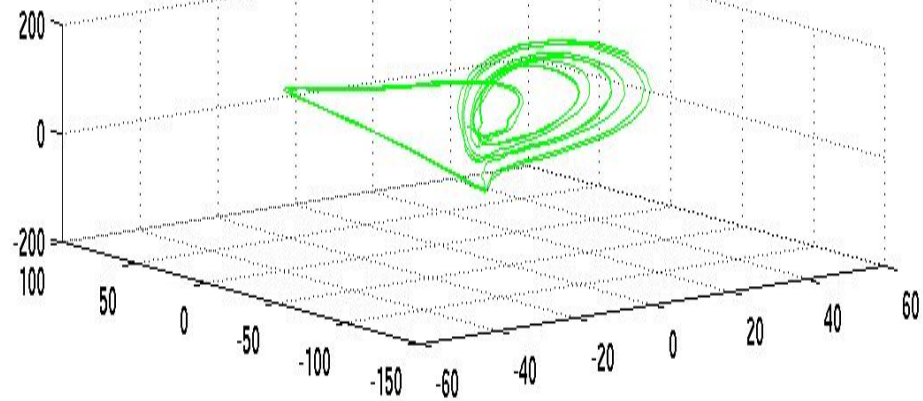


Extra-tropical Atmosphere



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

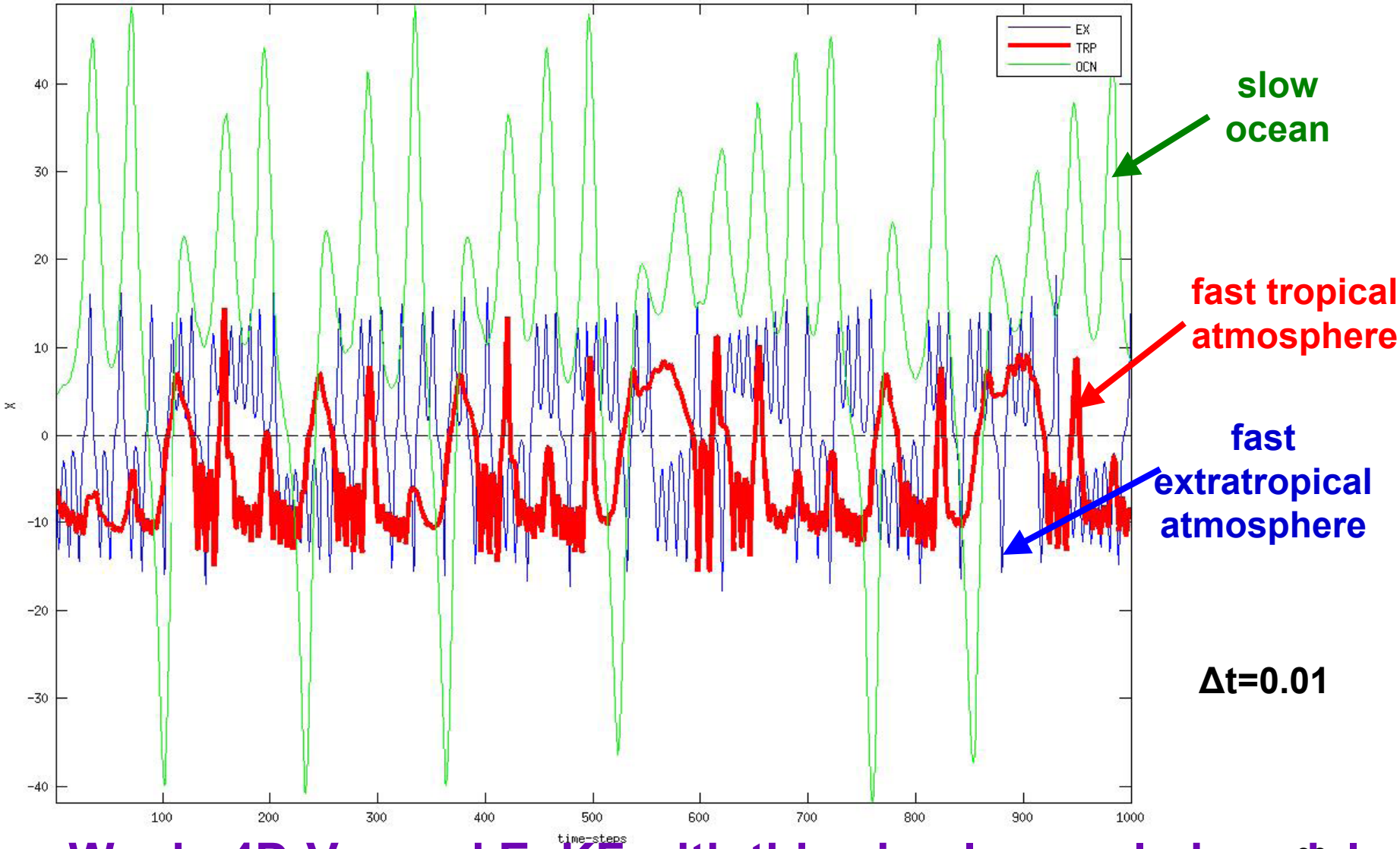
Tropical Ocean



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

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

Time series of the x-component



We do 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

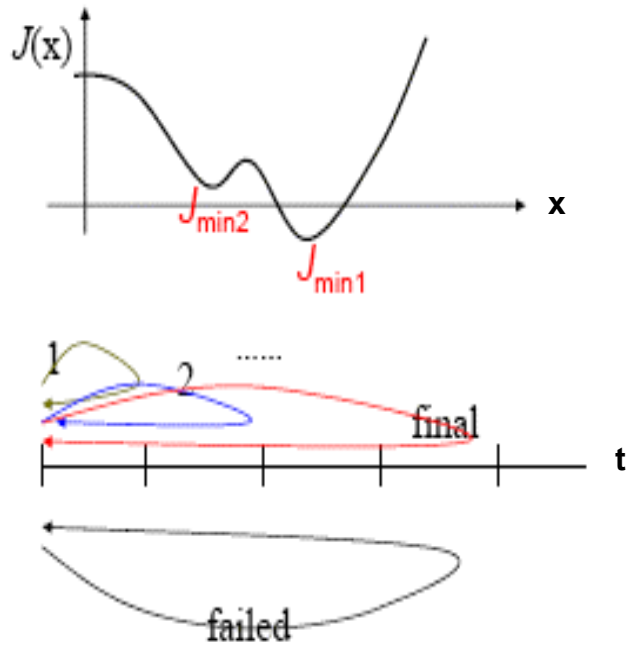
4D-Var: Quasi-static Variational Data Assimilation (QVA)

- For longer windows, multiple minima are a problem for 4D-Var minimization (Pires et al., 1996).

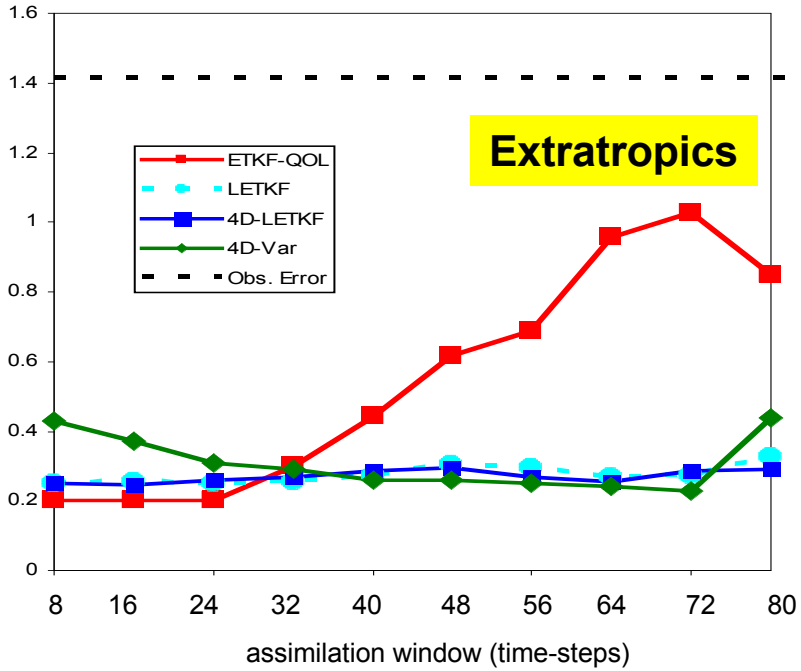
- Also for longer assimilation windows, non-Gaussian perturbations of the observation error and background error -> in non-quadratic cost functions

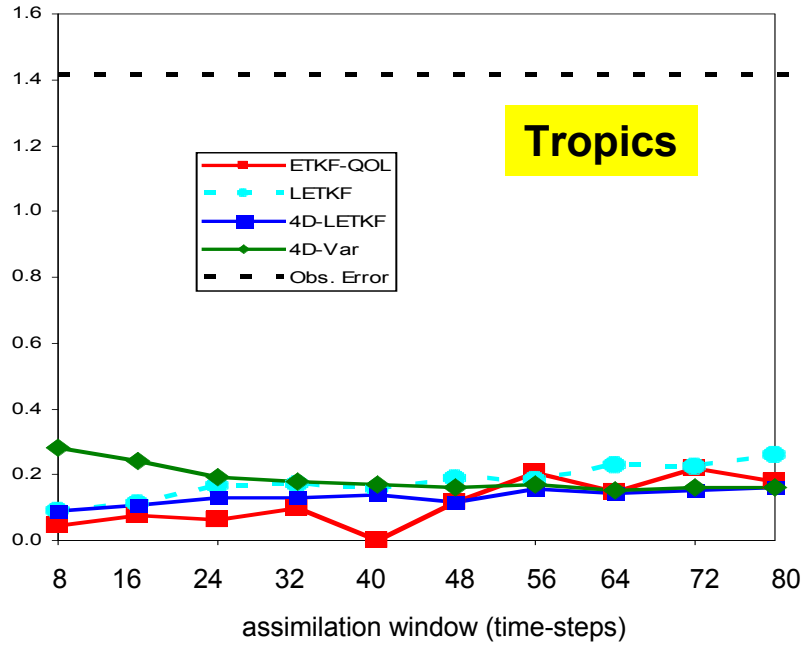
- Pires et al. (1996) proposed the Quasi-static Variational Data Assimilation (QVA) approach.

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

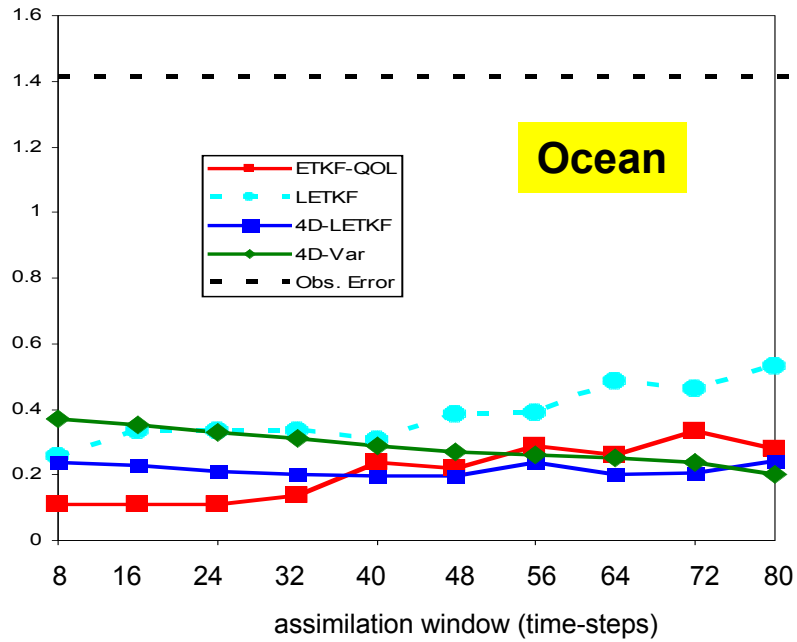


Fully coupled 4D-Var and EnKF: shorter windows



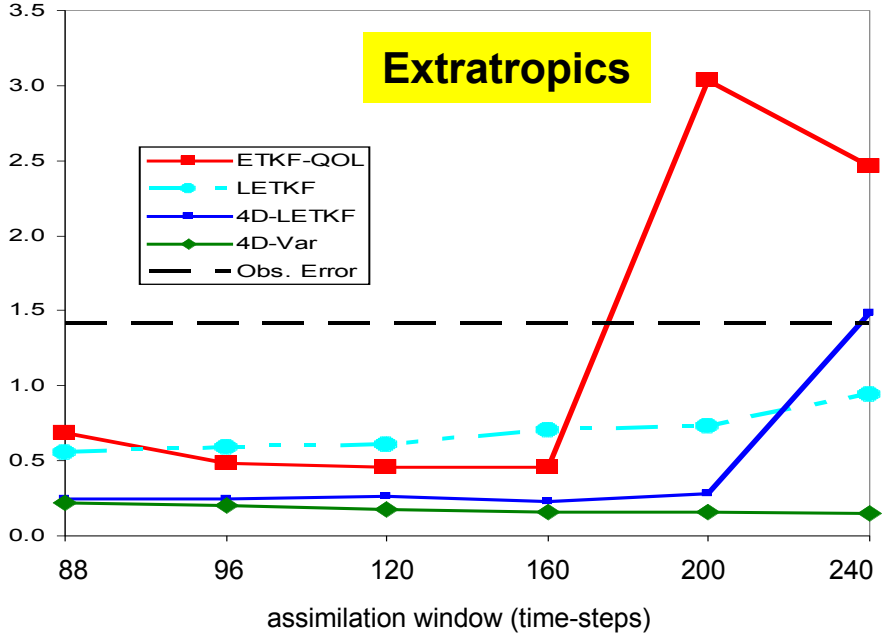


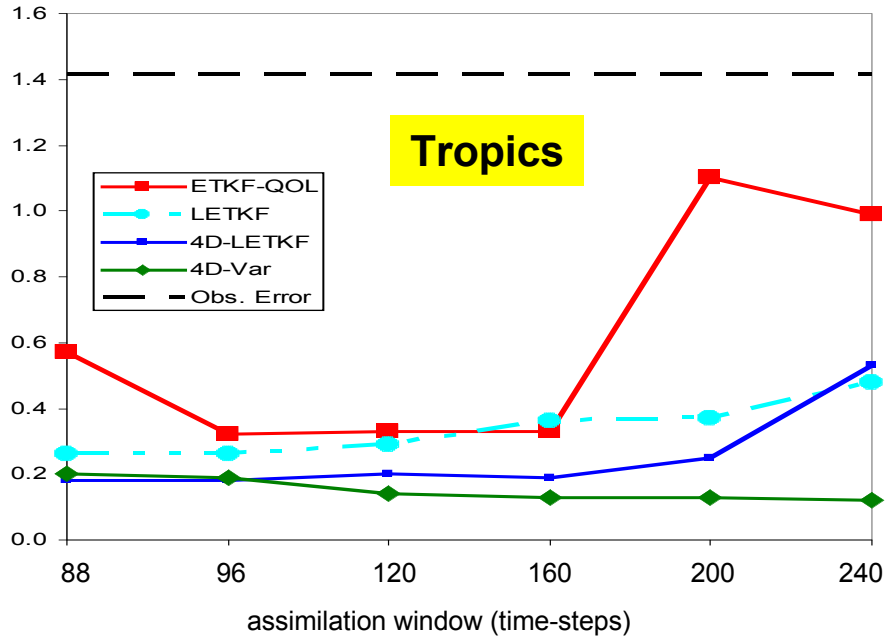
ETKF-QOL provides the best analysis for very short windows



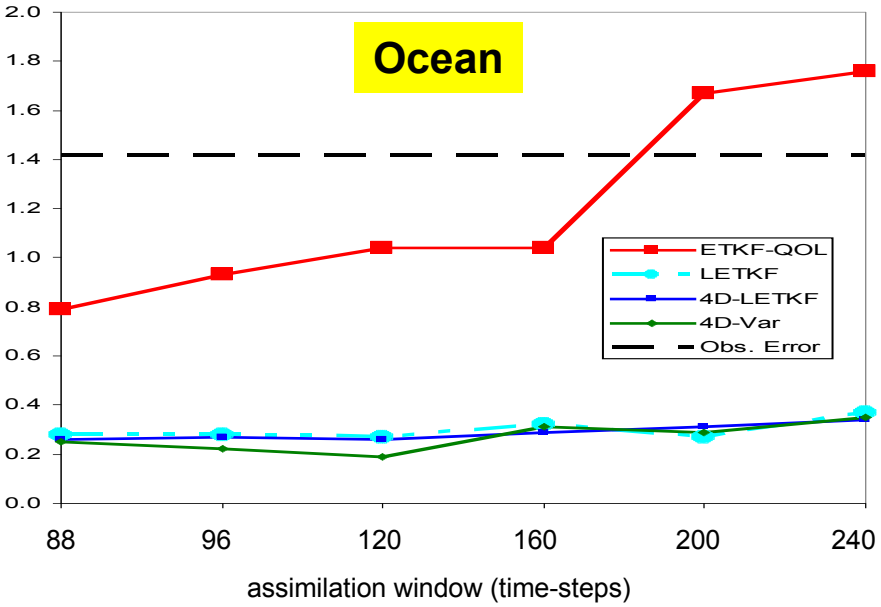
4D-Var competes with EnKF-based methods for longer windows

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.**

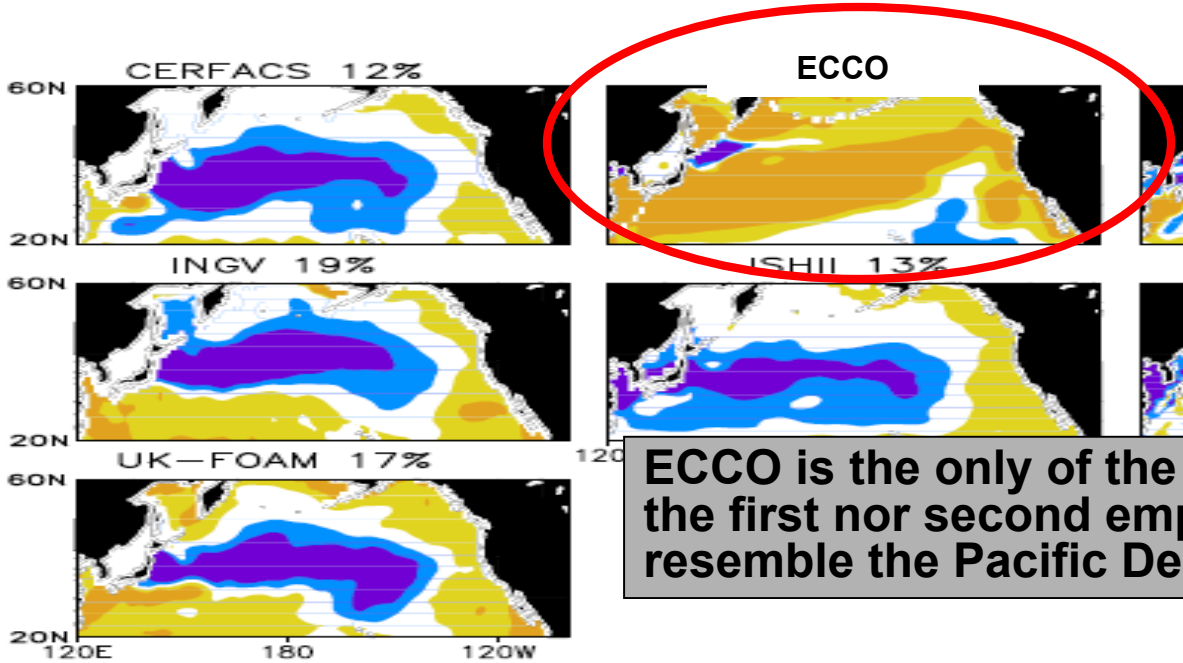
ECCO-like 4D-Var

- 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

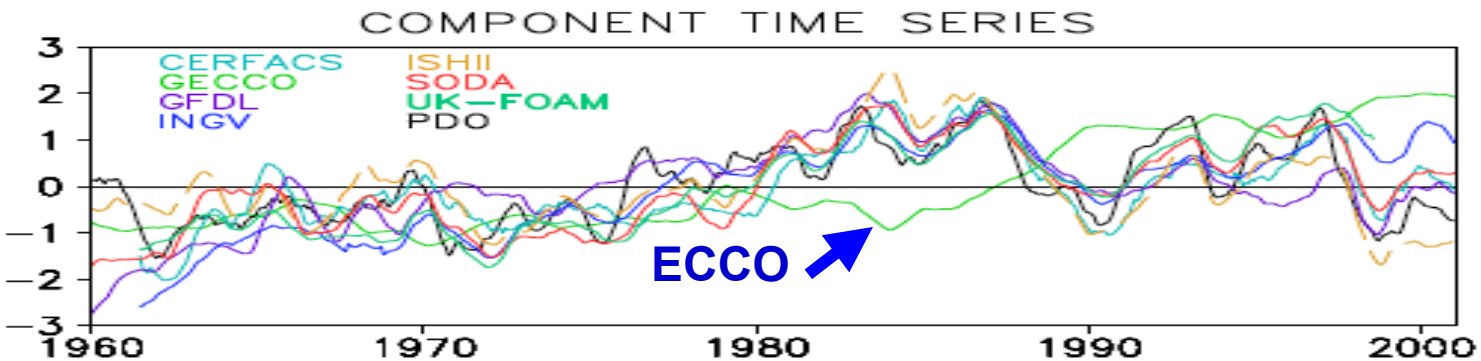
- The main characteristic of ECCO is that they include surface fluxes as control variables.
  - This allows them to have exceedingly long assimilation windows in 4D-Var (e.g. 10 years or even 50 years).
  - They used NCEP Reanalysis fluxes (Kalnay et al, 1996) as a first guess.

- 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 of the analyses for which neither the first nor second empirical eigenfunction 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: The Cost Function

The cost function to be minimized is given by

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

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

where

$$\mathbf{x}_{0,f} = \left(X_0, Y_0, Z_0 \quad f_1^1, f_2^1, f_3^1 \quad f_1^2, f_2^2, f_3^2 \quad \ldots \quad f_1^n, f_1^n, f_1^n \right)^T$$

Initial model state

Fluxes for first 8 time steps

Fluxes for last 8 time steps

$$\mathbf{x}^{b,nfe} = \left(X^b, Y^b, Z^b \quad f_1^{nfe,1}, f_2^{nfe,1}, f_3^{nfe,1} \quad f_1^{nfe,2}, f_2^{nfe,2}, f_3^{nfe2} \quad \ldots \quad f_1^{nfe,n}, f_1^{nfe,n}, f_1^{nfe,n} \right)^T$$

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

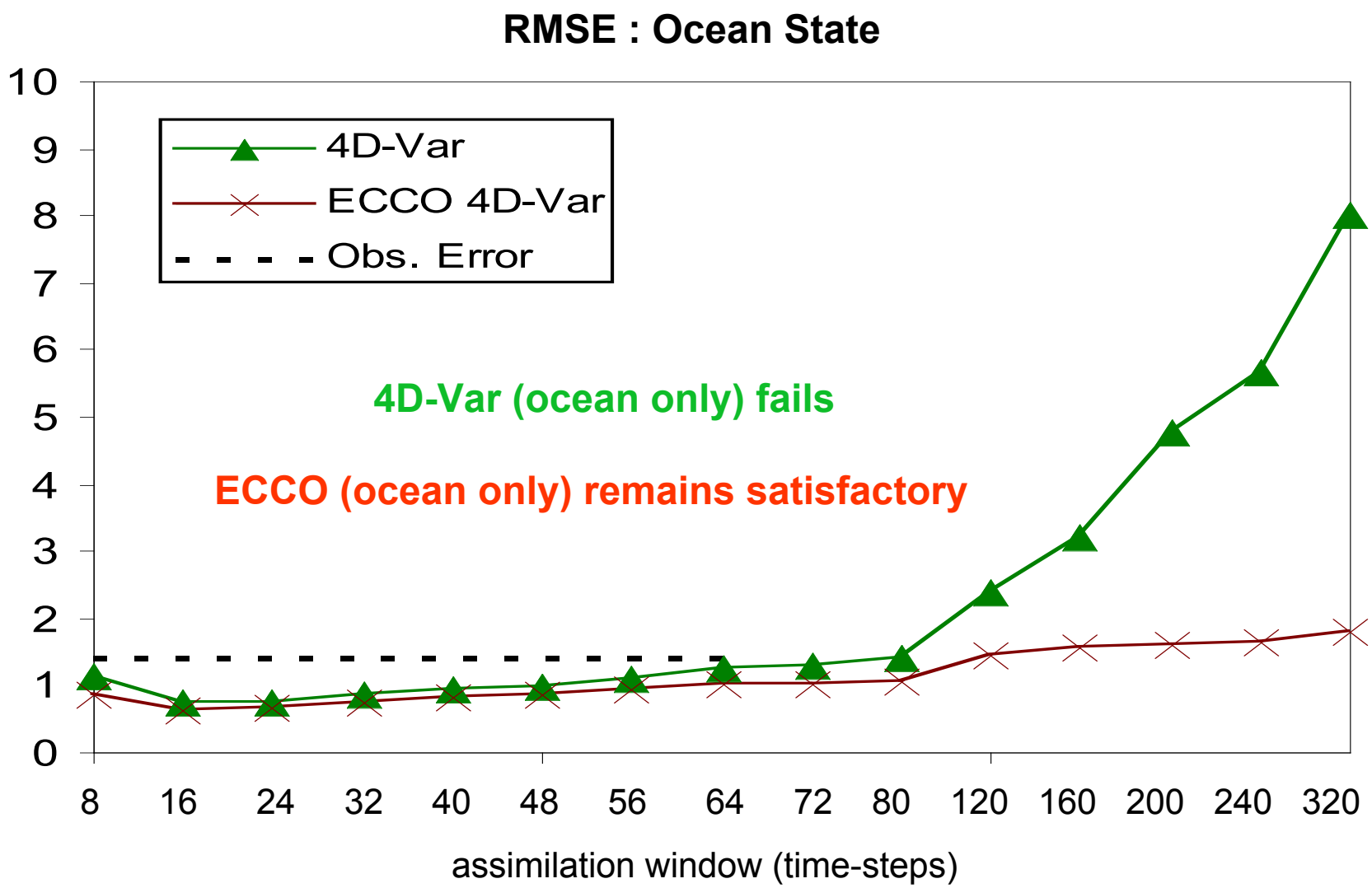
- **Observations**
  - Same as EnKF and 4D-Var experiments
- **Forecast Model**
  - Slow subsystem of coupled model with fluxes changing after every 8 time steps
- **Data Assimilation: ECCO Ocean 4D-Var**
  - **Control variables are <u>initial ocean state and flux terms</u>**
  - Prescribed background 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

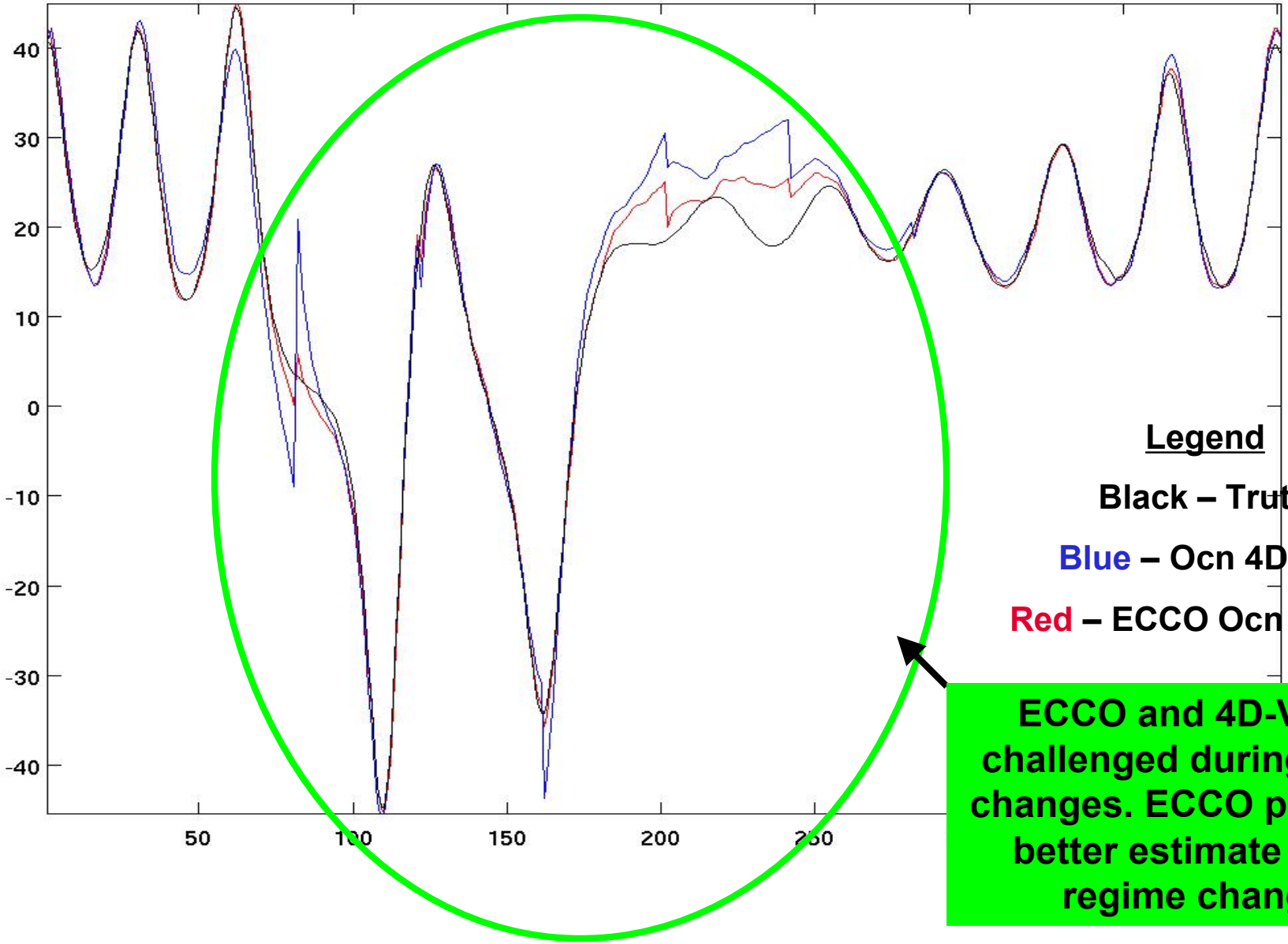
Comparison of ECCO-like & Ocean 4D-Var

QVA APPLIED

OCEAN ONLY

Obs. s.d. error = 1.41 for ocean





Legend

Black – Truth

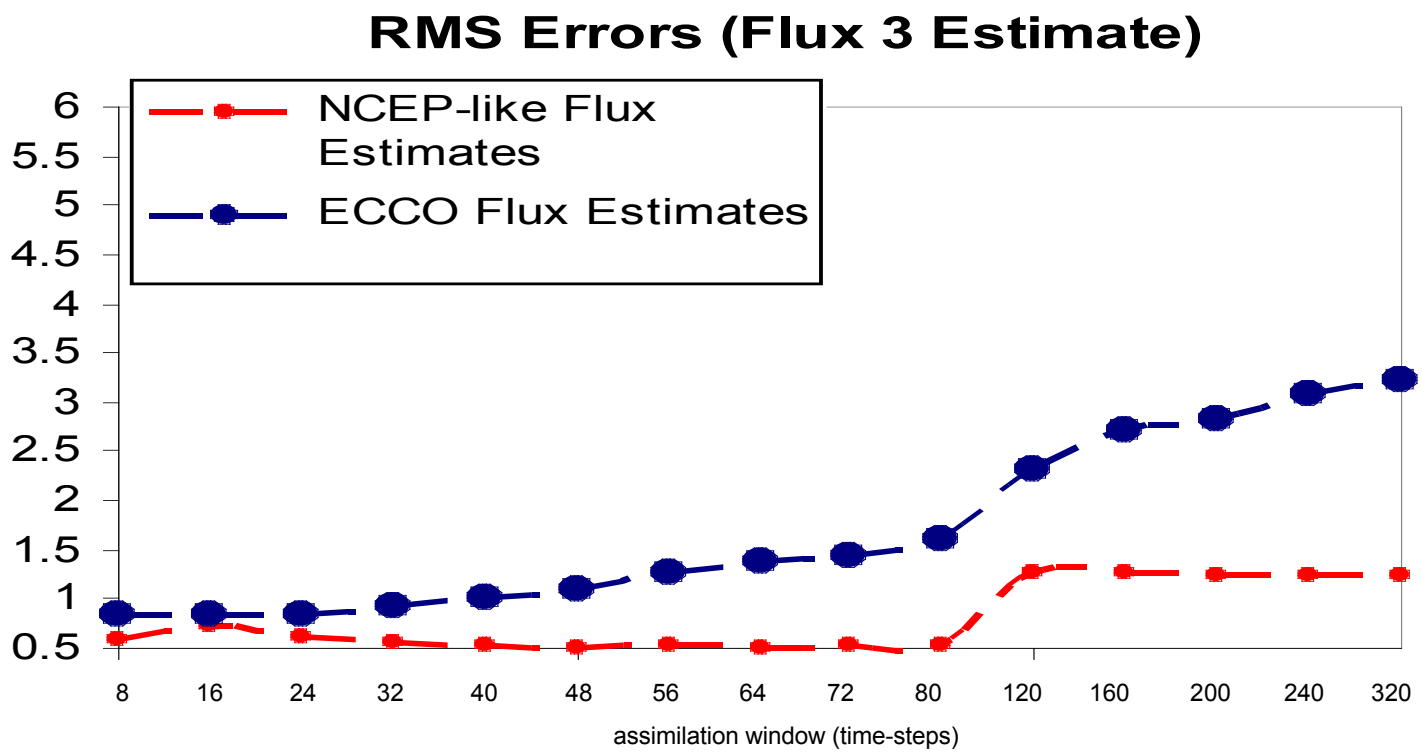
Blue – Ocn 4D-Var

Red – ECCO Ocn 4D-Var

ECCO and 4D-Var are challenged during regime changes. ECCO provides a better estimate during regime changes.

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, 2010, 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