EnKF, 4D-Var and ECCO in a "toy" ocean-atmosphere model

Tamara Singleton[1,2], Eugenia Kalnay[1], Kayo Ide[1] and Shu-Chih Yang[3]

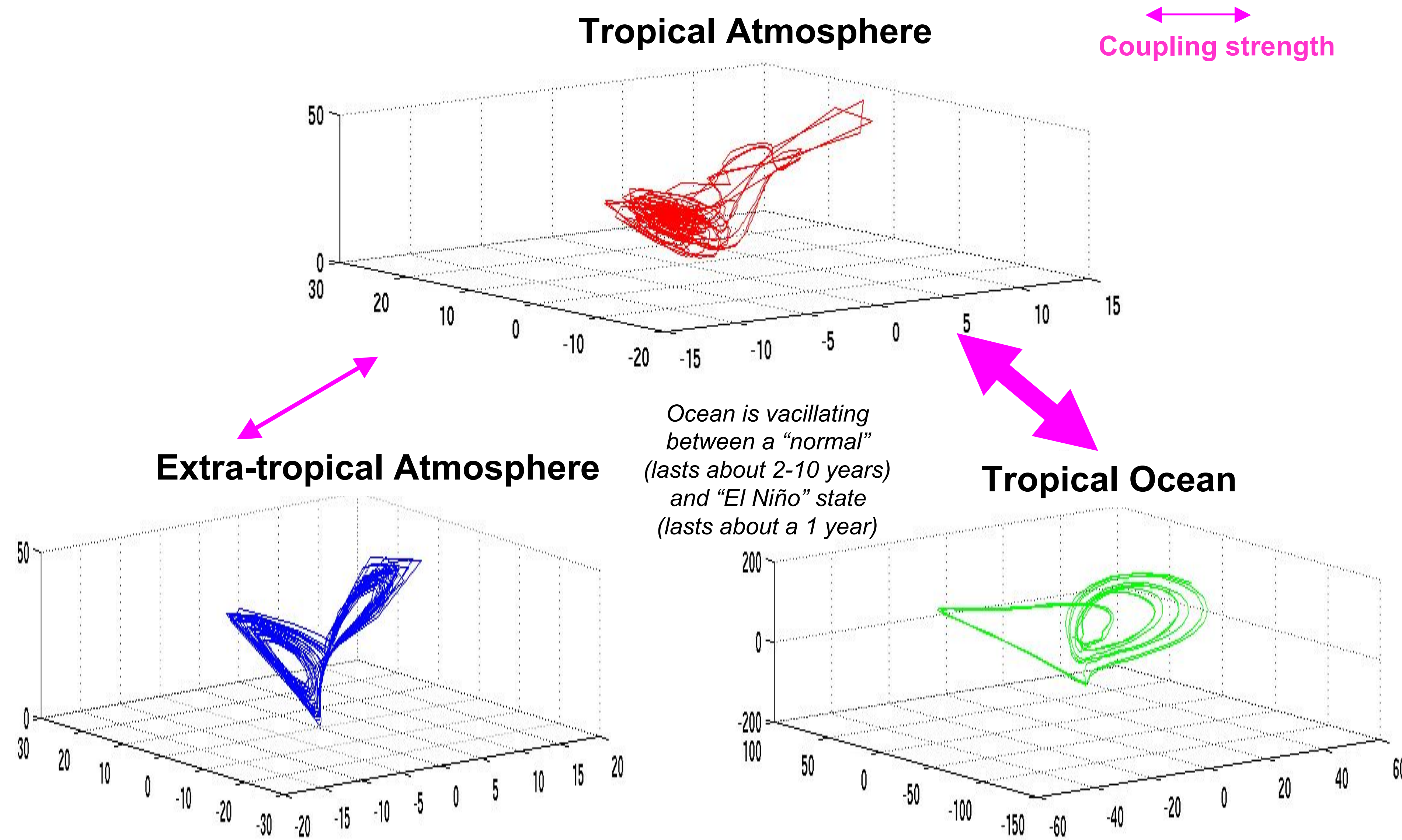
[1]UMD, [2]Johns Hopkins, [3]Taiwan NCU

EnKF, 4D-Var and ECCO in a "toy" ocean-atmosphere model

Tamara Singleton[1,2], Eugenia Kalnay[1], Kayo Ide[1] and Shu-Chih Yang[3]

[1]UMD, [2]Johns Hopkins, [3]Taiwan NCU

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



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

Questions explored:

-- Which is more accurate: 4D-Var or EnKF?
-- Should we use short or long windows?
-- Is it better to do an ocean reanalysis separately, or as a single coupled system?
-- Should we use frequent atmospheric observations in a coupled system?
-- Would RIP/QOL be beneficial in a coupled system?

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

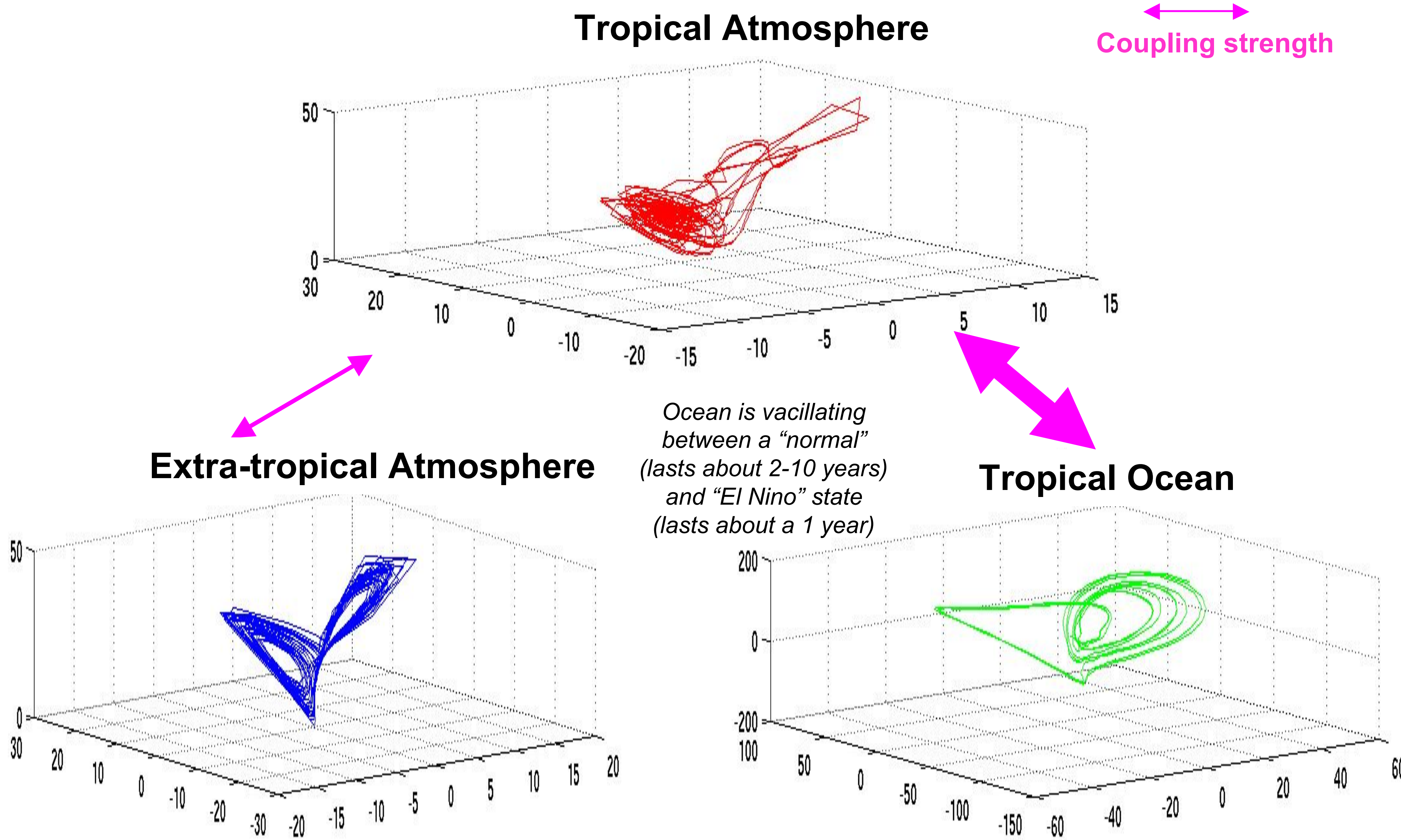
$$\dot{x}_t = \sigma(y_t - x_t) - c(X + k_2) - c_e(x_e + k_1)$$
$$\dot{y}_t = rx_t - y_t - x_t z_t + c(Y + k_2) + c_e(y_e + k_1)$$
$$\dot{z}_t = x_t y_t - bz_e + c_z Z$$

Ocean

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

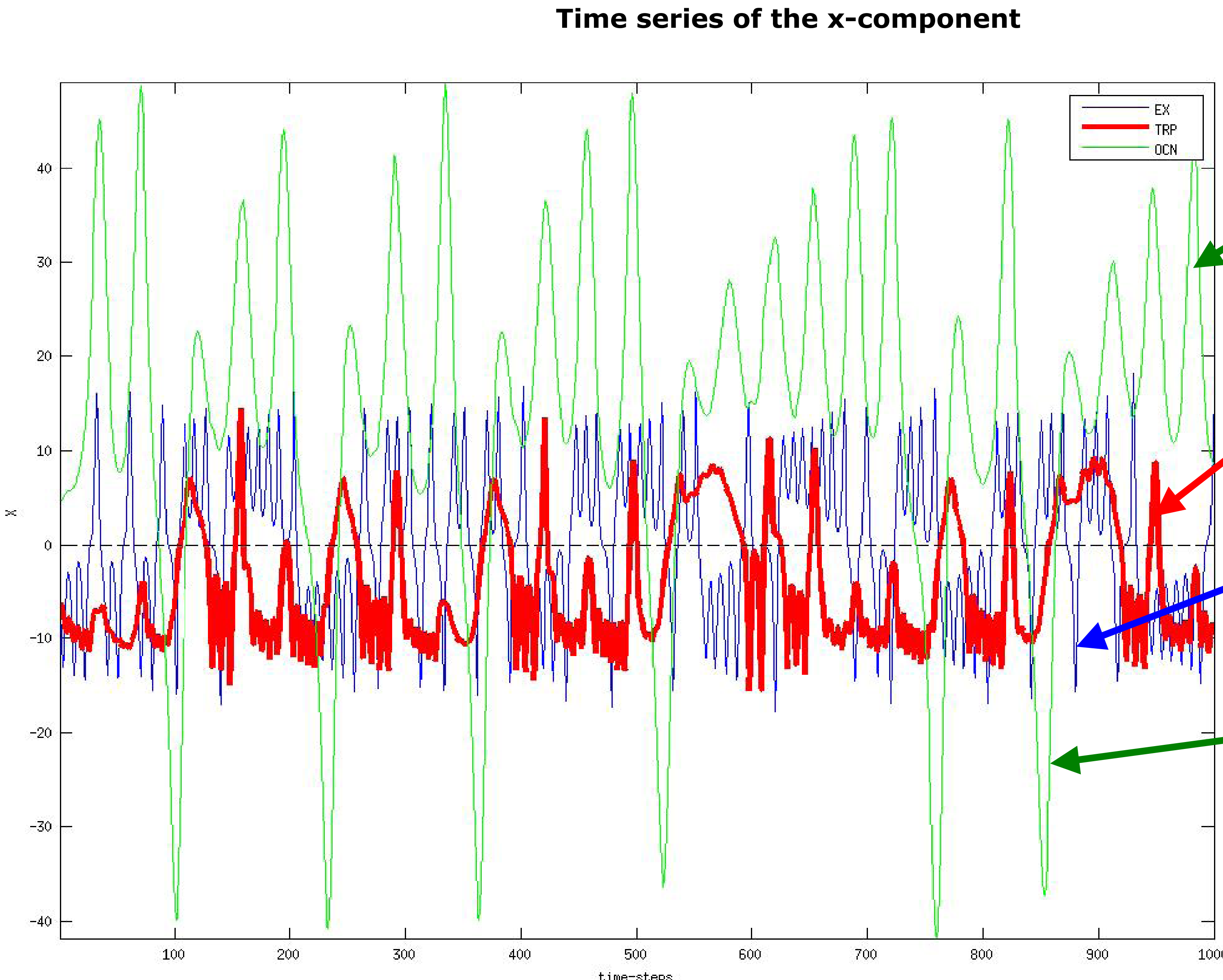
Model State: $[x_e, y_e, z_e, x_t, y_t, z_t, X, Y, Z]^T$

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



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

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



slow ocean "normal years"

fast tropical atmosphere

fast extratropical atmosphere

slow ocean "El Niño years"

$\Delta t=0.01$

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 $\sqrt{2}$ 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

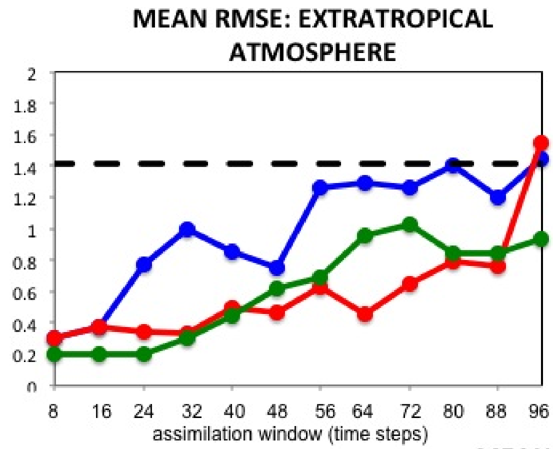
Description of EnKF-based methods

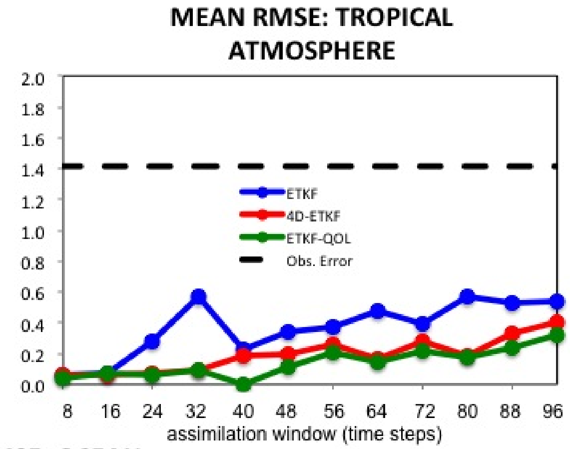
Method	Assimilating	Observations	Special Features
ETKF (Fully coupled)	Fast and slow variables simultaneously	Available at the end of a window (analysis time)	
4D-ETKF (Fully coupled)	Fast and slow variables simultaneously	Available throughout an assimilation window	4-dimensional
ETKF-QOL (Fully coupled)	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

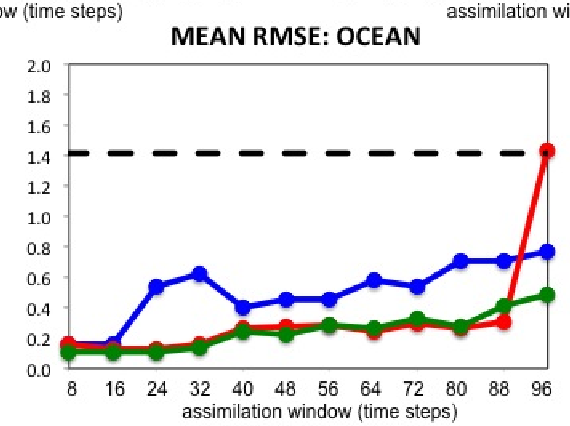
Coupled ocean-atmosphere ensembles:

ETKF, 4D-ETKF, ETKF-QOL

RMS error as a function of assimilation window length







The fully coupled ETKF data assimilations work well.
The shortest assimilation window (8 steps) is the best.
4D-ETKF (assimilating all the obs) is better than ETKF for longer windows. ETKF-QOL has the best performance (short windows).

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

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

J^b - "background" cost function

J^o - "observation" cost function

where the control variables are the initial 9 model variables:

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

Initial model state for extratropical atmos

Initial model state for tropical atmos.

Initial model state for ocean

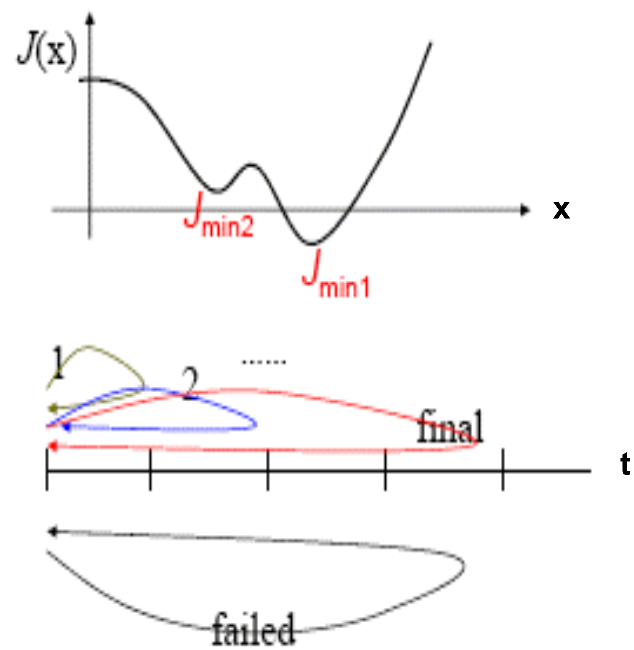
4D-Var: Quasi-static Variational Data Assimilation (QVA)

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

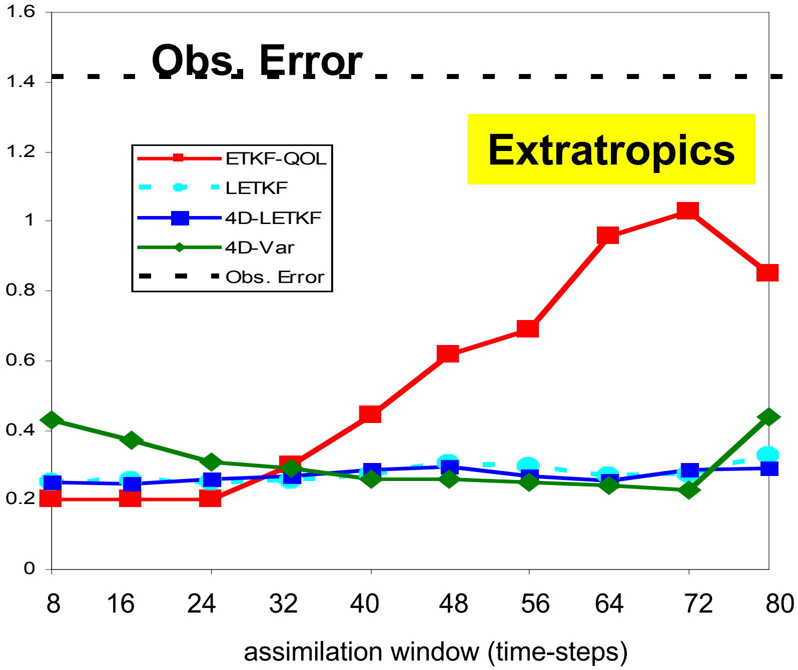
- Also for longer assimilation windows, non-Gaussian perturbations of the observation error and background error -> in non-quadratic cost functions

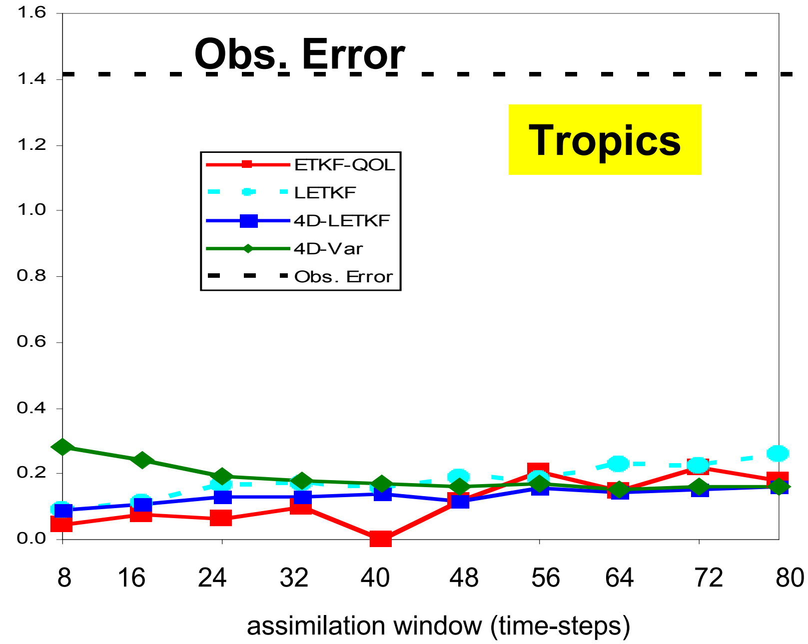
- Pires et al. (1996) proposed the Quasi-static Variational Data Assimilation (QVA) approach.

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

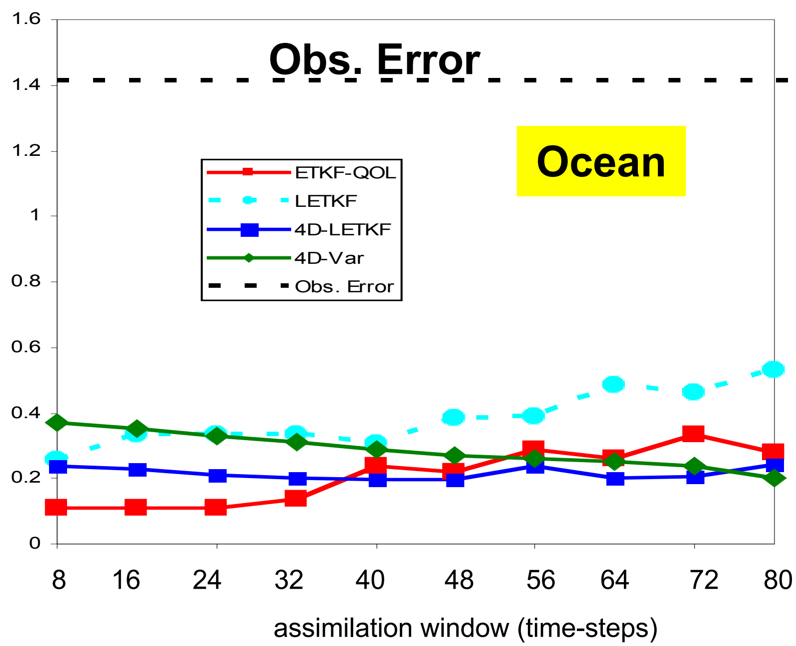


Fully coupled 4D-Var (+QVA) and EnKF: shorter windows



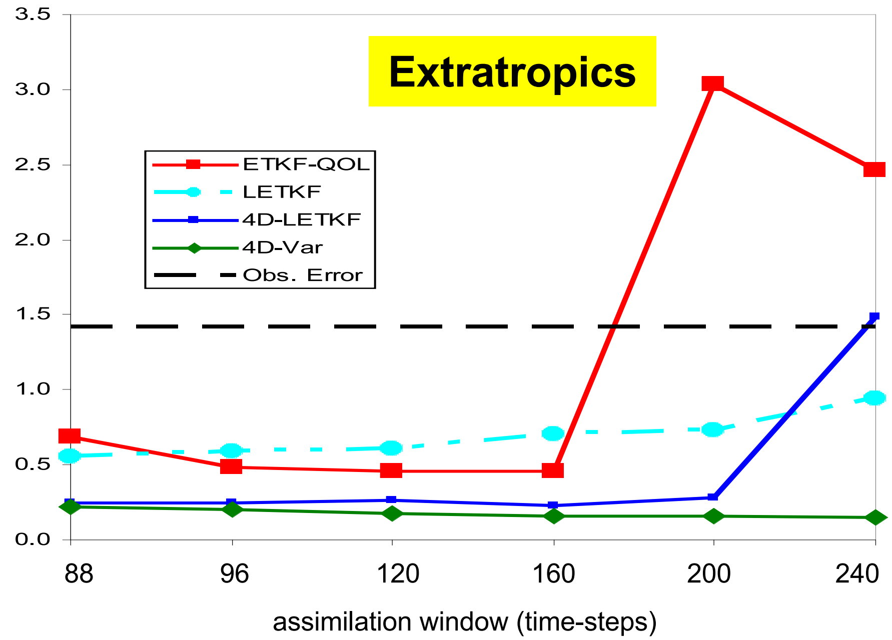


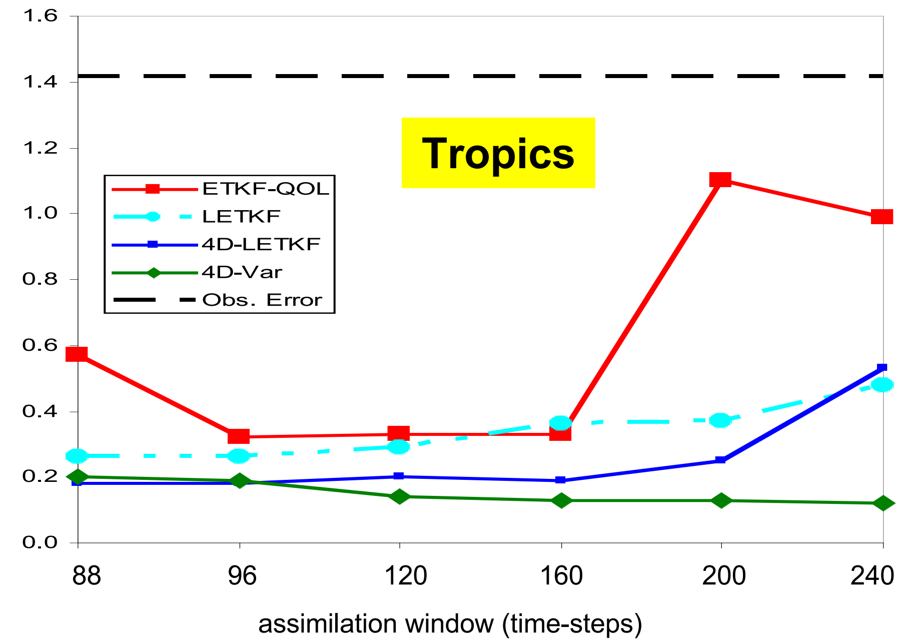
ETKF-QOL provides the best analysis for very short windows



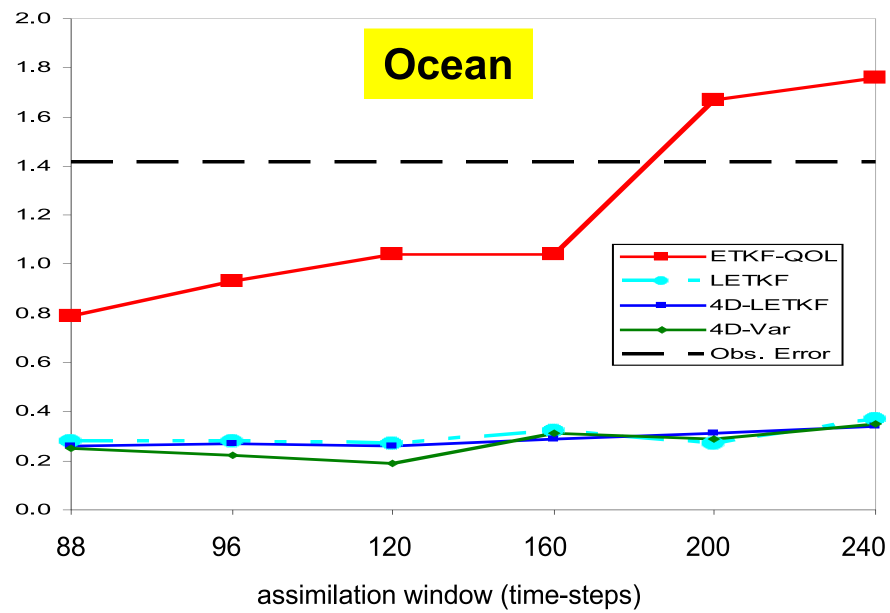
4D-Var competes with EnKF-based methods for longer windows

Fully coupled 4D-Var (+QVA) 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
- Lengthening the assimilation windows and applying QVA improves the 4D-Var analysis because 4D-Var “forgets” B. But longer windows are more expensive…
- Fully coupled EnKF are optimal for short windows. Short windows are less expensive…
- EnKF+QOL works best (short windows).
- **The optimal configurations (short windows for EnKF and long windows for 4D-Var) have similar accuracy.**

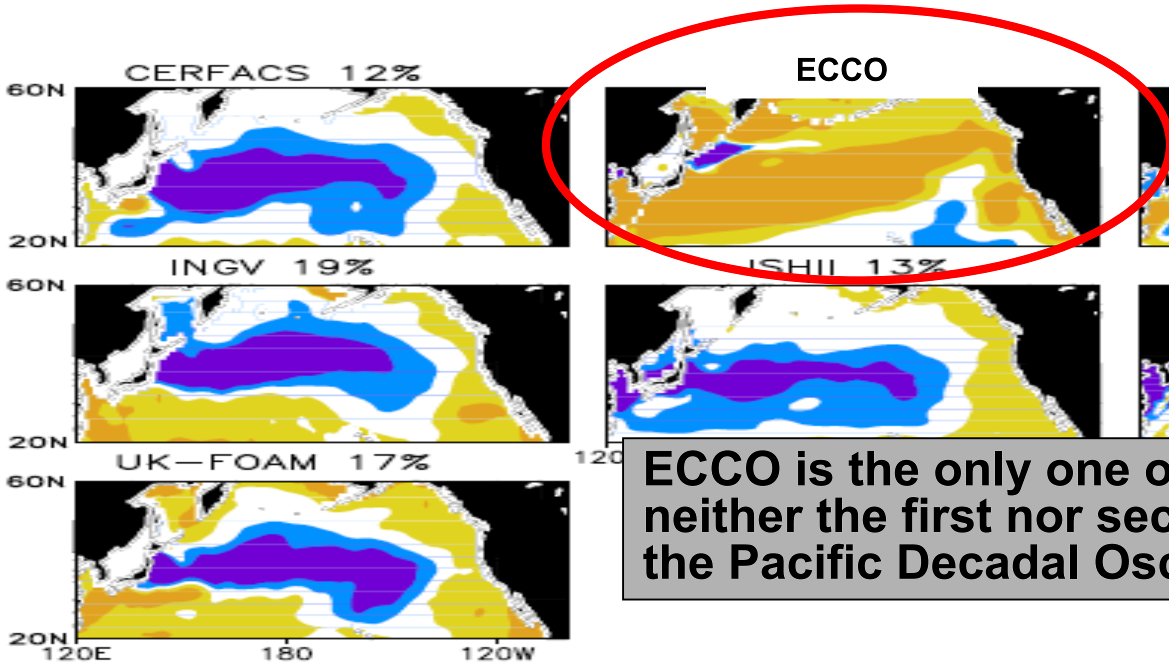
ECCO-like 4D-Var

- 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

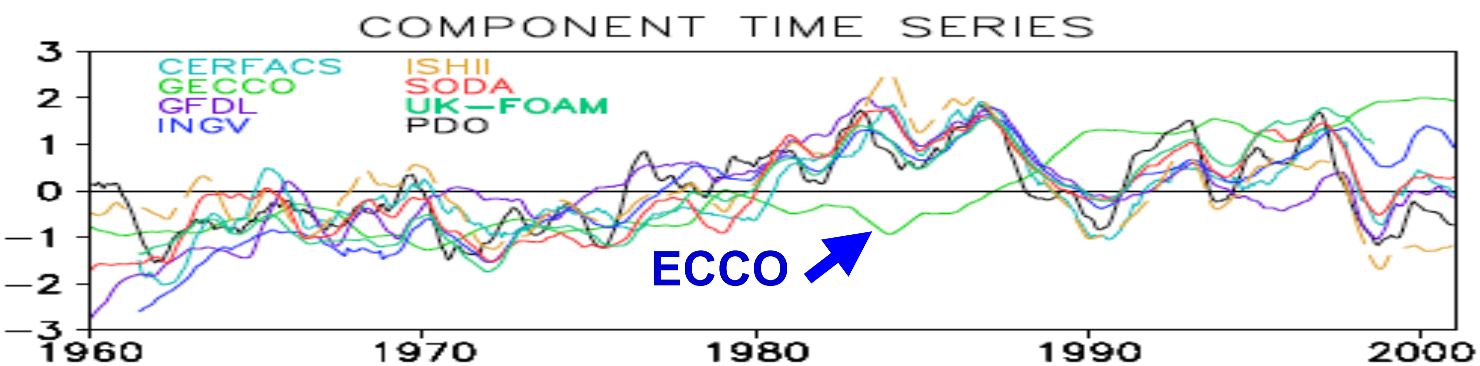
- The main characteristic of ECCO is that they include surface fluxes as control variables.
  - This allows them to have exceedingly long assimilation windows in 4D-Var (e.g. 10 years or even 50 years).
  - They used NCEP Reanalysis fluxes (Kalnay et al, 1996) as a first guess.

- 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

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

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

where the control variables are:

$$\mathbf{x}_{0,f} = \Big(\underbrace{X_0, Y_0, Z_0}_{\text{Initial model state}} \;\; \underbrace{f_1^1, f_2^1, f_3^1}_{\text{Fluxes for first 8 time steps}} \;\; f_1^2, f_2^2, f_3^2 \;\; \ldots \;\; \underbrace{f_1^n, f_1^n, f_1^n}_{\text{Fluxes for last 8 time steps}} \Big)^T$$

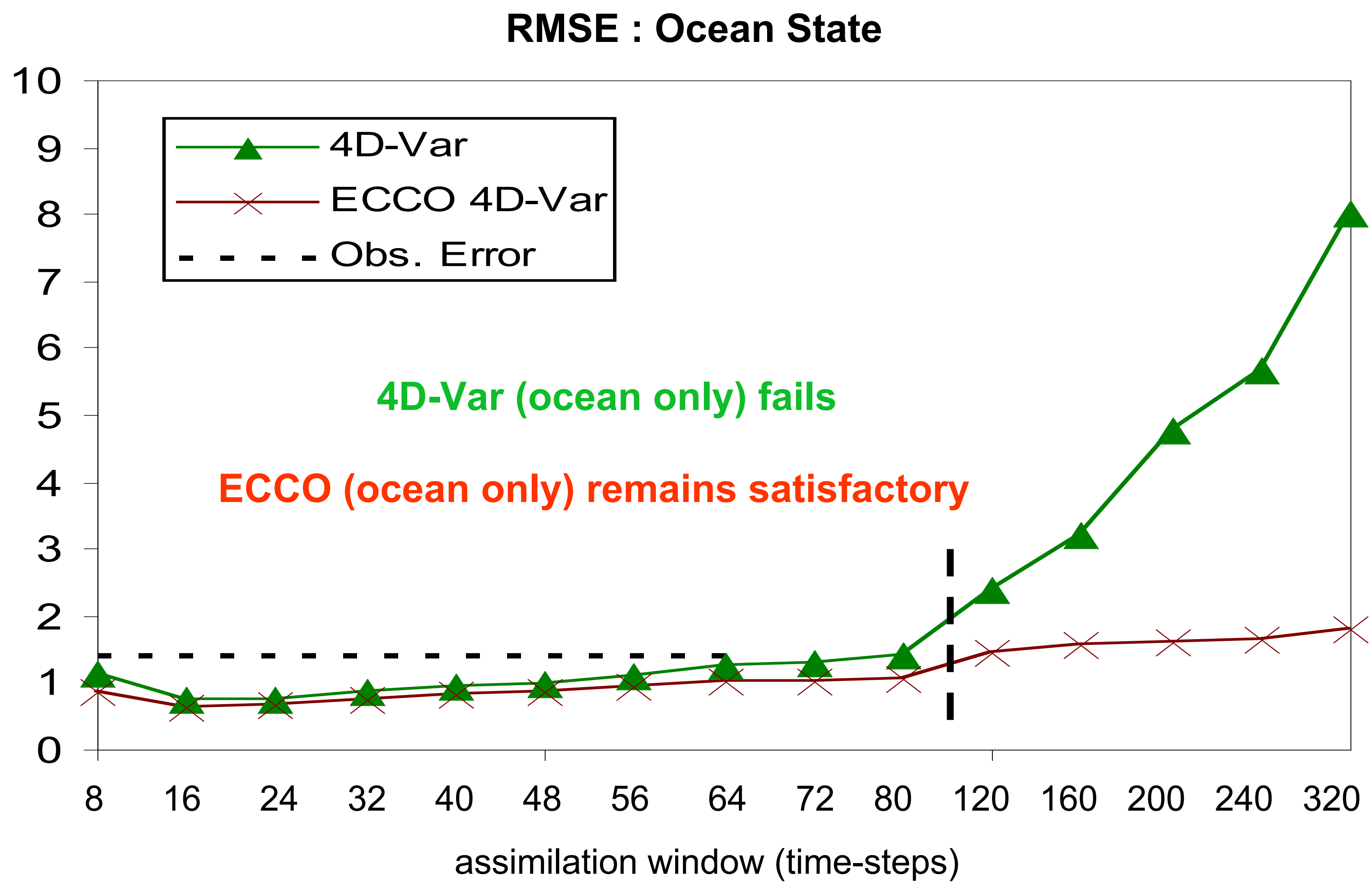
$$\mathbf{x}^{b,nfe} = \Big(\underbrace{X^b, Y^b, Z^b}_{\text{Background state for the ocean}} \;\; \underbrace{f_1^{nfe,1}, f_2^{nfe,1}, f_3^{nfe,1}}_{\text{NCEP-like flux estimates for first 8 time steps}} \;\; f_1^{nfe,2}, f_2^{nfe,2}, f_3^{nfe2} \;\; \ldots \;\; \underbrace{f_1^{nfe,n}, f_1^{nfe,n}, f_1^{nfe,n}}_{\text{NCEP-like flux estimates for last 8 time steps}} \Big)^T$$

Comparison of ECCO-like & Ocean 4D-Var

QVA APPLIED

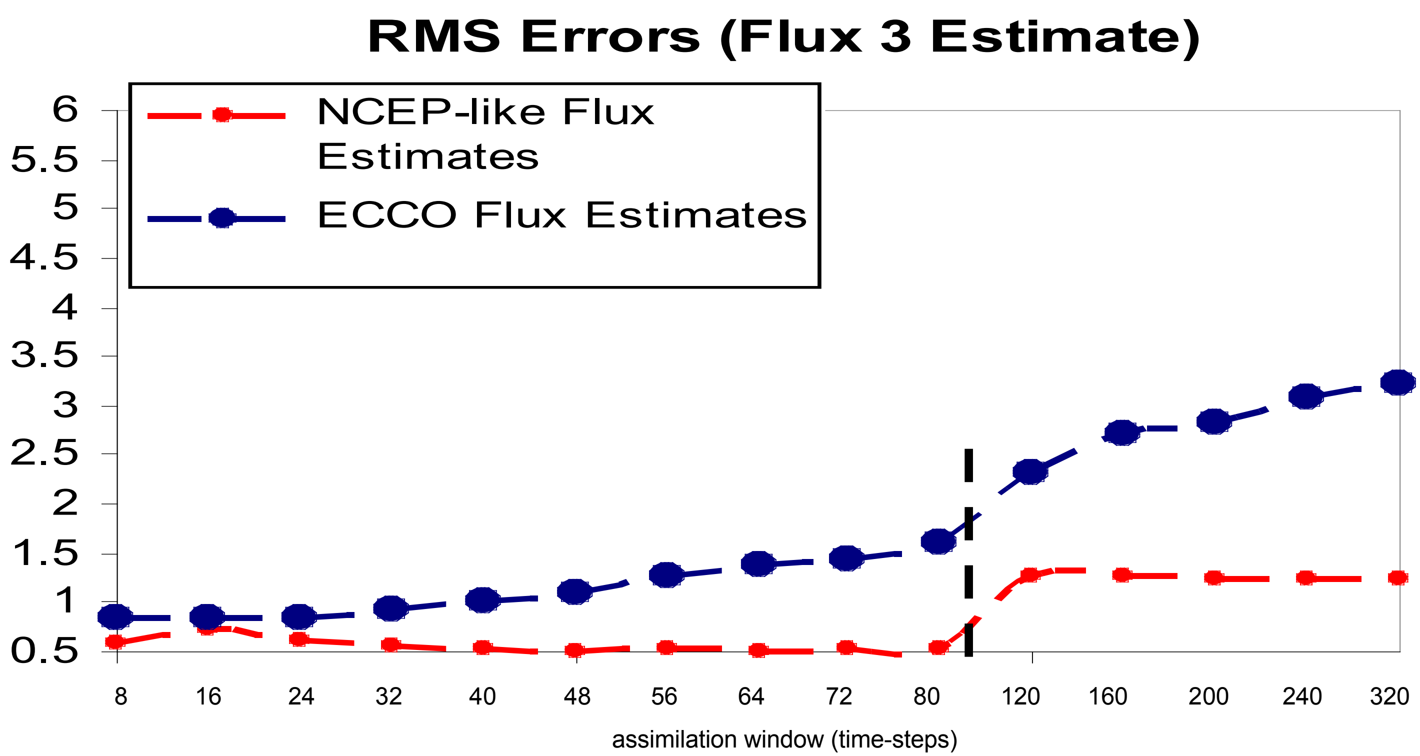
OCEAN ONLY

Obs. s.d. error = 1.41 for ocean



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

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. Both can handle frequent atmospheric observations.

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. Both can handle frequent atmospheric observations.

-- 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. Both can handle frequent atmospheric observations.

-- 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.

-- 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.

Practical Conclusions for Coupled Assimilation

- Since EnKF is as accurate as 4D-Var, and is optimal for short windows, it is more efficient than 4D-Var, which requires very long windows for maximum accuracy.

- Contrary to our expectations, the best results included frequent atmospheric observations.

- ECCO is 4D-Var including surface fluxes as control variables. This allows ECCO to have very long windows (decades).

- Since the estimated surface fluxes “adapt” in order to force the ocean model to be close to the observations, they are not guaranteed to be more accurate than the background fluxes.

- In our toy coupled model, the estimated surface fluxes were less accurate than the background fluxes.

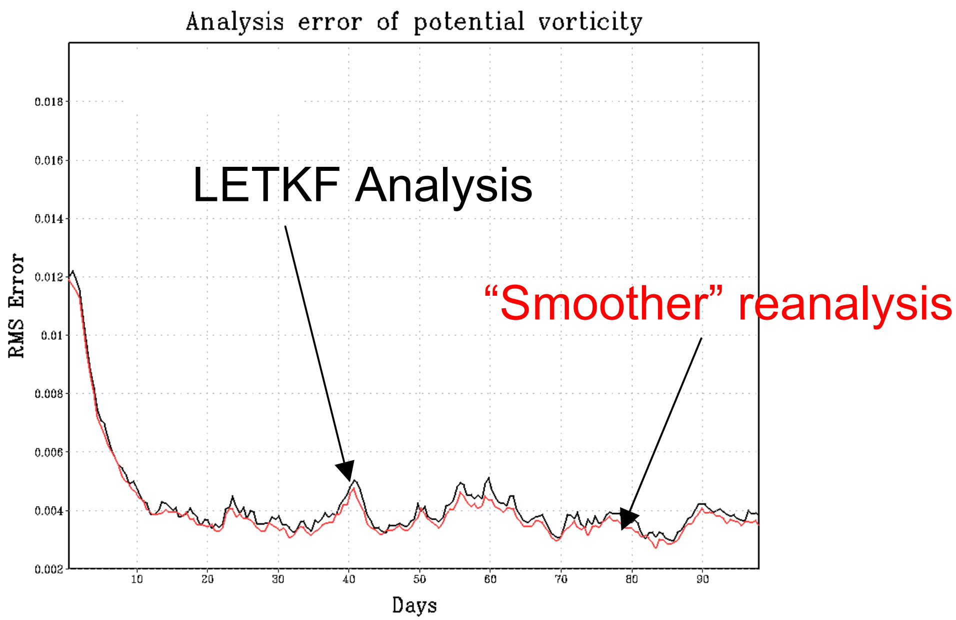
Extra: no-cost LETKF smoother allows a comparison of EnKF initial and final increments: the initial 4D-Var increments are sensitive to the norm, the final increments are similar to EnKF

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)