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

Tamara Singleton\textsuperscript{1,2}, Eugenia Kalnay\textsuperscript{1}, Kayo Ide\textsuperscript{1} and Shu-Chih Yang\textsuperscript{3}

\textsuperscript{1}UMD, \textsuperscript{2}Johns Hopkins, \textsuperscript{3}Taiwan NCU
EnKF, 4D-Var and ECCO in a “toy” ocean-atmosphere model

Tamara Singleton\textsuperscript{1,2}, Eugenia Kalnay\textsuperscript{1}, Kayo Ide\textsuperscript{1} and Shu-Chih Yang\textsuperscript{3}

\textsuperscript{1}UMD, \textsuperscript{2}Johns Hopkins, \textsuperscript{3}Taiwan NCU

Coupling strength

Tropical Atmosphere

Ocean is vacillating between a “normal” (lasts about 2-10 years) and “El Niño” state (lasts about a 1 year)

Extra-tropical Atmosphere

Tropical Ocean

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?
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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>c, c(_z), c(_e)</td>
<td>Coupling coefficient</td>
<td>c, c(_z) = 1, c(_e) = 0.08</td>
</tr>
<tr>
<td>(\tau)</td>
<td>time scale</td>
<td>(\tau = 0.1)</td>
</tr>
<tr>
<td>(\sigma), (b), and (r)</td>
<td>Lorenz parameters</td>
<td>(\sigma=10), (b=8/3), and (r=28)</td>
</tr>
<tr>
<td>(k_1, k_2)</td>
<td>Uncentering parameters</td>
<td>(k_1=10), (k_2 = -11)</td>
</tr>
</tbody>
</table>

Extratropical atmosphere
\[
\begin{align*}
\dot{x}_e &= \sigma(y_e - x_e) - c_e(x_t + k_1) \\
\dot{y}_e &= rx_e - y_e - x_e z_e - c_e(y_t + k_1) \\
\dot{z}_e &= x_e y_e - b z_e
\end{align*}
\]

Tropical atmosphere
\[
\begin{align*}
\dot{x}_t &= \sigma(y_t - x_t) - c(X + k_2) - c_e(x_e + k_1) \\
\dot{y}_t &= rx_t - y_t - x_t z_t + c(Y + k_2) + c_e(y_e + k_1) \\
\dot{z}_t &= x_t y_t - b z_e + c_z Z
\end{align*}
\]

Ocean
\[
\begin{align*}
\dot{X} &= \tau \sigma(Y - X) - c(x_t + k_2) \\
\dot{Y} &= \tau r X - \tau Y - \tau X Z + c(y_t + k_2) \\
\dot{Z} &= \tau X Y - \tau b Z + c_z z_t
\end{align*}
\]

Model State: \([x_e, y_e, z_e, x_t, y_t, z_t, X, Y, Z]^T\)
We compare 4D-Var and EnKF with this simple coupled model
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

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

\[
J(x_{t_0}) = \frac{1}{2} \left[ x_{t_0} - x_{t_0}^b \right]^T B_0^{-1} \left[ x_{t_0} - x_{t_0}^b \right] + \frac{1}{2} \sum_{i=1}^N \left[ H(x_{t_i}) - y_{t_i}^o \right]^T R_{t_i}^{-1} \left[ H(x_{t_i}) - y_{t_i}^o \right]
\]

where the control variables are the initial 9 model variables:

\[
x_0 = \left( x_e^0, y_e^0, z_e^0, x_t^0, y_t^0, z_t^0, X^0, Y^0, Z^0 \right)^T
\]

- \(J^b\) - "background" cost function
- \(J^o\) - "observation" cost function

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.
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

Extratropics

Tropics

Ocean

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.
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.

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

\[
J = \frac{1}{2} [x_{0,f} - x^{b,nfe}]^T (B^{0,nfe})^{-1} [x_{0,f} - x^{b,nfe}] + \frac{1}{2} \sum_{i=1}^{N} [Hx_{t_i} - y^o_{t_i}]^T (R_{t_i}^{-1}) [Hx_{t_i} - y^o_{t_i}]
\]

where the control variables are:

\[
x_{0,f} = \begin{pmatrix} X_0, Y_0, Z_0 \ f_1^1, f_2^1, f_3^1 \ f_1^2, f_2^2, f_3^2 \ \vdots \ f_n^1, f_1^n, f_1^n \end{pmatrix}^T
\]

\[
x^{b,nfe} = \begin{pmatrix} X^b, Y^b, Z^b \ f_{nfe,1}^1, f_{nfe,1}^2, f_{nfe,1}^3 \ f_{nfe,2}^1, f_{nfe,2}^2, f_{nfe,2}^3 \ \vdots \ f_{nfe,n}^1, f_{nfe,n}^1, f_{nfe,n}^1 \end{pmatrix}^T
\]

Initial model state
Fluxes for first 8 time steps
Background state for the ocean
NCEP-like flux estimates for first 8 time steps

Fluxes for last 8 time steps
NCEP-like flux estimates for last 8 time steps

B^{0,nfe} = \begin{pmatrix} B & 0 \\ 0 & Q \end{pmatrix}
Comparison of ECCO-like & Ocean 4D-Var

QVA APPLIED  OCEAN ONLY  Obs. s.d. error = 1.41 for ocean

RMSE : Ocean State

4D-Var (ocean only) fails
ECCO (ocean only) remains satisfactory

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.
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.
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.

\[ \bar{x}^a_n = \bar{x}^f_n + X^f_n \bar{w}_n \]

\[ \hat{x}^a_{n-1} = \bar{x}^f_{n-1} + X^f_{n-1} \bar{w}_n \]

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)