Ensemble Strategies for State and Parameters Estimation in Ocean Ecosystem Models

– Joint, Dual, and OSA-based EnKF schemes –

Workshop on Meteorological Sensitivity Analysis and Data Assimilation
Roanoke, West Virginia

M.E. Gharamti*†, L. Bertino*, B. Ait-El-Fquih‡, A. Samuelsen*, I. Hoteit‡

* Nansen Environmental and Remote Sensing Center, Bergen, NORWAY
‡ King Abdullah University of Science and Technology, Thuwal, KSA

† E-mail: mohamad.gharamti@nersc.no
Webpage: http://www.nersc.no/group/ecosystem-modeling-group

June, 2015
Context: Ocean Ecosystem Modeling

Coupled models

- *NorESM*: Norwegian Earth System Model; coupled atm-land-ice-ocean(MICOM)-biogeochemistry(HAMOCC)
- *TOPAZ-ECO*: physics(HYCOM,GOTM)-biology(ECOSMO,NORWECOM)
**Context: Ocean Ecosystem Modeling**

**Coupled models**

- *NorESM*: Norwegian Earth System Model; coupled atm-land-ice-ocean(MICOM)-biogeochemistry(HAMOCC)
- *TOPAZ-ECO*: physics(HYCOM,GOTM)-biology(ECOSMO,NORWECOM)

**Biological data**

- Satellite: surface chlorophyll-a
- In-situ: Nutrients concentrations, pCO₂, ..
Context: Ocean Ecosystem Modeling

Coupled models

- **NorESM**: Norwegian Earth System Model; coupled atm-land-ice-ocean(MICOM)-biogeochemistry(HAMOCC)
- **TOPAZ-ECO**: physics(HYCOM,GOTM)-biology(ECOSMO,NORWECOM)

Biological data

- Satellite: surface chlorophyll-a
- In-situ: Nutrients concentrations, pCO$_2$, ..

DA framework and usage

- Combined state-parameters estimation (EnKF)
- Dimension, non-linearties (bloom), complexity
  - Environmental monitoring – Fisheries
  - Initialization for climate projections
Outline of the Talk

Problem statement

Standard DA techniques

Alternative formulation of the state-parameters estimation problem

Application using a 1D ecosystem model

Conclusion
Challenges and Motivation

**Objective** ⇒ Need to find $p(x_k, \theta_k \mid y_{0:k})$ recursively in time using an EnKF
Challenges and Motivation

Objective $\Rightarrow$ Need to find $p(x_k, \theta_k|y_{0:k})$ recursively in time using an EnKF

Issues of the joint-filtering problem:

- Positive variables (concentration of nutrients, ...)
- Poorly known parameters (e.g., grazing efficiency)
- Noisy, seasonal (sparse) data extracted from “sub-optimal” locations!
Challenges and Motivation

**Objective** ⇒ Need to find $p(x_k, \theta_k | y_{0:k})$ recursively in time using an EnKF

**Issues of the joint-filtering problem:**

- Positive variables (concentration of nutrients, ...)
- Poorly known parameters (e.g., grazing efficiency)
- Noisy, seasonal (sparse) data extracted from “sub-optimal” locations!

**Gaussian anamorphosis:**
Transform both variables and observations (Bertino et al., 2003; Simon and Bertino 2009, 2012; Song et al., 2014)
**Challenges and Motivation**

**Objective** ⇒ Need to find \( p(x_k, \theta_k \mid y_{0:k}) \) recursively in time using an EnKF

**Issues of the joint-filtering problem:**

- Positive variables (concentration of nutrients, ...)
- Poorly known parameters (e.g., grazing efficiency)
- Noisy, seasonal (sparse) data extracted from “sub-optimal” locations!

**Gaussian anamorphosis:**
Transform both variables and observations (Bertino et al., 2003; Simon and Bertino 2009, 2012; Song et al., 2014)

**A simple alternative:** Truncate any negative variables after the analysis

- Possible depletion of the components of the ensemble “pdf deformation”
- Parameters updated in wrong directions (spring bloom time)
**Challenges and Motivation**

**Objective** ⇒ Need to find $p(x_k, \theta_k | y_{0:k})$ recursively in time using an EnKF

**Issues of the joint-filtering problem:**

- Positive variables (concentration of nutrients, ...)
- Poorly known parameters (e.g., grazing efficiency)
- Noisy, seasonal (sparse) data extracted from “sub-optimal” locations!

**Gaussian anamorphosis:**
Transform both variables and observations (Bertino et al., 2003; Simon and Bertino 2009, 2012; Song et al., 2014)

**A simple alternative:** Truncate any negative variables after the analysis

- Possible depletion of the components of the ensemble “pdf deformation”
- Parameters updated in wrong directions (spring bloom time)

**Our Approach:** Use simple truncation and propose a different and a more consistent formulation of the state-parameters estimation problem.
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form \( p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k}) \).

\[ \sim Update \text{ both the state and parameters simultaneously:} \]
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form $p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k})$.

$\sim Update$ both the state and parameters simultaneously:

$$p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \equiv p(z_{k-1} | y_{0:k-1})$$
State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form \( p(x_k, \theta_k|y_{0:k}) \rightarrow p(z_k|y_{0:k}) \).

\( \sim \) Update both the state and parameters simultaneously:

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) &= p(z_{k-1}|y_{0:k-1}) \\
p(x_k, \theta_k|y_{0:k-1}) &= p(z_k|y_{0:k-1})
\end{align*}
\]
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form $p(x_k, \theta_k|y_{0:k}) \rightarrow p(z_k|y_{0:k})$.

$\sim$ *Update both the state and parameters simultaneously:*

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) & \equiv p(z_{k-1}|y_{0:k-1}) \\
p(x_k, \theta_k|y_{0:k-1}) & \equiv p(z_k|y_{0:k-1}) \\
p(x_k, \theta_k|y_{0:k}) & \equiv p(z_k|y_{0:k})
\end{align*}
\]
State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p(\mathbf{x}_k, \theta_k | y_{0:k}) \rightarrow p(\mathbf{z}_k | y_{0:k})$.

$\leadsto$ Update both the state and parameters simultaneously:

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)
State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k})$.

$\rightarrow$ Update both the state and parameters simultaneously:

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) & \overset{F}{\rightarrow} p(x_k, \theta_k | y_{0:k-1}) \\
\equiv p(z_{k-1} | y_{0:k-1}) & \overset{A}{\rightarrow} p(x_k, \theta_k | y_{0:k}) \\
\quad \equiv p(z_k | y_{0:k})
\end{align*}
\]

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

Dual-EnKF: Separate the densities $p(x_k, \theta_k | y_{0:k}) \rightarrow p(\theta_k | y_{0:k}) \cdot p(x_k | \theta_k, y_{0:k})$.

$\rightarrow$ Update the parameters before the state:
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form \( p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k}) \).

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \equiv p(z_{k-1} | y_{0:k-1}) \]

\[ \mathcal{F} \rightarrow p(x_k, \theta_k | y_{0:k-1}) \equiv p(z_k | y_{0:k-1}) \]

\[ \mathcal{A} \rightarrow p(x_k, \theta_k | y_{0:k}) \equiv p(z_k | y_{0:k}) \]

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

**Dual-EnKF:** Separate the densities \( p(x_k, \theta_k | y_{0:k}) \rightarrow p(\theta_k | y_{0:k}) \cdot p(x_k | \theta_k, y_{0:k}) \).

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \]

\[ \sim Update the parameters before the state: \]
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form \( p(x_k, \theta_k|y_{0:k}) \rightarrow p(z_k|y_{0:k}) \).

\( \sim \) *Update both the state and parameters simultaneously:*

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) & \rightarrow \mathcal{F} \rightarrow \quad p(x_k, \theta_k|y_{0:k-1}) \\
& \equiv p(z_{k-1}|y_{0:k-1}) & \equiv p(z_k|y_{0:k-1}) \\
& \rightarrow \mathcal{A} \rightarrow \quad p(x_k, \theta_k|y_{0:k}) \\
& \equiv p(z_k|y_{0:k})
\end{align*}
\]

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

**Dual-EnKF:** Separate the densities \( p(x_k, \theta_k|y_{0:k}) \rightarrow p(\theta_k|y_{0:k}) \cdot p(x_k|\theta_k, y_{0:k}) \).

\( \sim \) *Update the parameters before the state:*

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) & \rightarrow \mathcal{A}^\theta \rightarrow \quad p(\theta_k|y_{0:k})
\end{align*}
\]
**State-Parameters Estimation (Standard Techniques)**

**Joint-EnKF:** Classical state-space augmented form $p(x_k, \theta_k|y_{0:k}) \rightarrow p(z_k|y_{0:k})$.

$\leadsto$ *Update both the state and parameters simultaneously:*

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) & \quad \triangleright \quad \mathcal{F} \quad \triangleright \quad p(x_k, \theta_k|y_{0:k-1}) \\
& \equiv p(z_{k-1}|y_{0:k-1}) & & \equiv p(z_k|y_{0:k-1}) \\
\mathcal{A} & \quad \triangleright \quad p(x_k, \theta_k|y_{0:k}) & & \equiv p(z_k|y_{0:k})
\end{align*}
\]

- Could yield to significant inconsistency \cite{Wen2006}
- Might be subject to stability and tractability issues \cite{Moradkhani2005, Wang2009}

**Dual-EnKF:** Separate the densities $p(x_k, \theta_k|y_{0:k}) \rightarrow p(\theta_k|y_{0:k}) \cdot p(x_k|\theta_k, y_{0:k})$.

$\leadsto$ *Update the parameters before the state:*

\[
\begin{align*}
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) & \quad \triangleright \quad \mathcal{A}^\theta \quad \triangleright \quad p(\theta_k|y_{0:k}) \\
\mathcal{F} & \quad \triangleright \quad p(x_k, \theta_k|y_{0:k})
\end{align*}
\]

5 / 19
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form \( p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k}) \).

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \equiv p(z_{k-1} | y_{0:k-1}) \]
\[ \mathcal{F} \]
\[ p(x_k, \theta_k | y_{0:k-1}) \equiv p(z_k | y_{0:k-1}) \]
\[ \mathcal{A} \]
\[ p(x_k, \theta_k | y_{0:k}) \equiv p(z_k | y_{0:k}) \]

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

**Dual-EnKF:** Separate the densities \( p(x_k, \theta_k | y_{0:k}) \rightarrow p(\theta_k | y_{0:k}) \cdot p(x_k | \theta_k, y_{0:k}) \).

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \]
\[ \mathcal{A}^\theta \]
\[ p(\theta_k | y_{0:k}) \]
\[ \mathcal{F} \]
\[ p(x_k | \theta_k, y_{0:k-1}) \]
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form $p(x_k, \theta_k | y_{0:k}) \rightarrow p(z_k | y_{0:k})$.

\[ \sim Update \text{ both the state and parameters simultaneously:} \]

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \quad \mathcal{F} \quad p(x_k, \theta_k | y_{0:k-1}) \quad \mathcal{A} \quad p(x_k, \theta_k | y_{0:k}) \]

- Could yield to significant inconsistency \((\text{Wen and Chen, 2006})\)
- Might be subject to stability and tractability issues \((\text{Moradkhani et al., 2005; Wang et al., 2009})\)

**Dual-EnKF:** Separate the densities $p(x_k, \theta_k | y_{0:k}) \rightarrow p(\theta_k | y_{0:k}) \cdot p(x_k | \theta_k, y_{0:k})$.

\[ \sim Update \text{ the parameters before the state:} \]

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \quad \mathcal{A}^0 \quad p(\theta_k | y_{0:k}) \quad \mathcal{F} \quad p(x_k | \theta_k, y_{0:k-1}) \]

\[ \mathcal{F} \quad p(x_k, \theta_k | y_{0:k}) \]
State-Parameters Estimation (Standard Techniques)

**Joint-EnKF:** Classical state-space augmented form \( p(x_k, \theta_k | y_{0:k}) \to p(z_k | y_{0:k}) \).

\[ \sim \text{Update both the state and parameters simultaneously:} \]

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \equiv p(z_{k-1} | y_{0:k-1}) \]
\[ \mathcal{F} \]
\[ p(x_k, \theta_k | y_{0:k-1}) \equiv p(z_k | y_{0:k-1}) \]
\[ A \]
\[ p(x_k, \theta_k | y_{0:k}) \equiv p(z_k | y_{0:k}) \]

- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

**Dual-EnKF:** Separate the densities \( p(x_k, \theta_k | y_{0:k}) \to p(\theta_k | y_{0:k}) \cdot p(x_k | \theta_k, y_{0:k}) \).

\[ \sim \text{Update the parameters before the state:} \]

\[ p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \]
\[ A^0 \]
\[ p(\theta_k | y_{0:k}) \]
\[ \mathcal{F} \]
\[ p(x_k | \theta_k, y_{0:k-1}) \]
\[ \mathcal{A}^x \]
\[ p(x_k, \theta_k | y_{0:k}) \]

- More expensive: requires 2 forward model integrations!
One-Step-Ahead Smoothing-based Joint-EnKF

**Problem statement**

**DA techniques**

**Alternative algorithmic formulation**

**Application: 1D ecosystem model**

**Summary**

One-Step-Ahead Smoothing-based Joint-EnKF

**Alternative formulation:**

- The classical path that involves the forecast pdf \( p(z_k|y_{0:k-1}) \) when moving from the analysis pdf \( p(z_{k-1}|y_{0:k-1}) \) to the analysis pdf at the next time \( p(z_k|y_{0:k}) \) is **not unique**!
One-Step-Ahead Smoothing-based Joint-EnKF

Alternative formulation:

- The classical path that involves the forecast pdf $p(z_k | y_{0:k-1})$ when moving from the analysis pdf $p(z_{k-1} | y_{0:k-1})$ to the analysis pdf at the next time $p(z_k | y_{0:k})$ is not unique!
- We resort to using the one-step-ahead smoothing pdf $p(z_{k-1} | y_{0:k})$. 
One-Step-Ahead Smoothing-based Joint-EnKF

Alternative formulation:

▶ The classical path that involves the forecast pdf $p(z_k|y_{0:k-1})$ when moving from the analysis pdf $p(z_{k-1}|y_{0:k-1})$ to the analysis pdf at the next time $p(z_k|y_{0:k})$ is not unique!

▶ We resort to using the one-step-ahead smoothing pdf $p(z_{k-1}|y_{0:k})$.

1. Smoothing Step: $p(x_{k-1}, \theta|y_{0:k})$ is first computed using likelihood $p(y_k|x_{k-1}, \theta)$:

$$p(x_{k-1}, \theta|y_{0:k}) \propto p(y_k|x_{k-1}, \theta) p(x_{k-1}, \theta|y_{0:k-1})$$
One-Step-Ahead Smoothing-based Joint-EnKF

Alternative formulation:

- The classical path that involves the forecast pdf \( p(z_k|y_{0:k-1}) \) when moving from the analysis pdf \( p(z_{k-1}|y_{0:k-1}) \) to the analysis pdf at the next time \( p(z_k|y_{0:k}) \) is not unique!
- We resort to using the one-step-ahead smoothing pdf \( p(z_{k-1}|y_{0:k}) \).

1. **Smoothing Step**: \( p(x_{k-1}, \theta|y_{0:k}) \) is first computed using likelihood \( p(y_k|x_{k-1}, \theta) \):

\[
p(x_{k-1}, \theta|y_{0:k}) \propto p(y_k|x_{k-1}, \theta)p(x_{k-1}, \theta|y_{0:k-1})
\]

2. **Analysis Step**: \( p(x_k|y_{0:k}) \) is computed using posteriori transition \( p(x_k|x_{k-1}, \theta, y_k) \):

\[
p(x_n|y_{0:k}) = \int p(x_k|x_{k-1}, \theta, y_k)p(x_{k-1}, \theta|y_{0:k})dx_{k-1}d\theta
\]
**Ensemble Implementation: Joint-EnKF_{OSA}**

**Joint-EnKF_{OSA}:** Reversed order of the forecast-update steps. \( \sim \) *Smooth the state and update the parameters before performing the forecast:*
Ensemble Implementation: Joint-EnKF_{OSA}

**Joint-EnKF_{OSA}:** Reversed order of the forecast-update steps. \(\sim\) *Smooth the state and update the parameters before performing the forecast:*

\[
p(x_{k-1}, \theta_{k-1} | y_{0:k-1})
\]
**Ensemble Implementation: Joint-EnKF\textsubscript{OSA}**

**Joint-EnKF\textsubscript{OSA}:** Reversed order of the forecast-update steps. ⚔️ *Smooth the state and update the parameters before performing the forecast:*

\[
p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \xrightarrow{\mathcal{A}, S^x} p(x_{k-1}, \theta_k | y_{0:k})
\]
**Ensemble Implementation: Joint-EnKF_{OSA}**

**Joint-EnKF_{OSA}:** Reversed order of the forecast-update steps. \(\leadsto Smooth\ the\ state\ and\ update\ the\ parameters\ before\ performing\ the\ forecast:\

\[
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) \xrightarrow{\mathcal{A}^\theta, S^x} p(x_{k-1}, \theta|y_{0:k})
\]
Ensemble Implementation: Joint-EnKF$_{OSA}$

**Joint-EnKF$_{OSA}$**: Reversed order of the forecast-update steps. \( \sim \) *Smooth the state and update the parameters before performing the forecast:*

\[
p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \xrightarrow{A^\theta, S^x} p(x_{k-1}, \theta_k | y_{0:k}) \xrightarrow{\mathcal{F}} p(x_k, \theta_k | y_{0:k})
\]
**Ensemble Implementation: Joint-EnKF\textsubscript{OSA}**

**Joint-EnKF\textsubscript{OSA}:** Reversed order of the forecast-update steps. \(\rightsquigarrow \) *Smooth the state and update the parameters before performing the forecast:*

\[
p(x_{k-1}, \theta_{k-1} | y_{0:k-1}) \xrightarrow{\mathcal{A}^\theta, S^x} p(x_{k-1}, \theta_{k} | y_{0:k}) \xrightarrow{\mathcal{F}} p(x_{k}, \theta_{k} | y_{0:k})
\]

---

**1. Smoothing Step**

\[
y^{f,(m)}_k = H_k \left( M_{k-1}(x^{a,(m)}_{k-1}, \theta^{(m)}_{|k-1}) + u^{(m)}_{k-1} \right) + v^{(m)}_k ; \quad v^{(m)}_k \sim \mathcal{N}(0, R_k)
\]

\[
x^{s,(m)}_{k-1} = x^{a,(m)}_{k-1} + P_{x^{a,(m)}_{k-1}, y^{f}_k} P^{-1}_{y^{f}_k} (y_k - y^{f,(m)}_k)
\]

\[
\theta^{(m)}_{|k} = \theta^{(m)}_{|k-1} + P_{\theta_{|k-1}, y^{f}_k} P^{-1}_{y^{f}_k} (y_k - y^{f,(m)}_k)
\]
**Ensemble Implementation: Joint-EnKF\textsubscript{OSA}**

**Joint-EnKF\textsubscript{OSA}:** Reversed order of the forecast-update steps. \(\sim\) *Smooth the state and update the parameters before performing the forecast:*

\[
p(x_{k-1}, \theta_{k-1}|y_{0:k-1}) \xrightarrow{\mathcal{A}^\theta, S^x} p(x_{k-1}, \theta|y_{0:k}) \xrightarrow{\mathcal{F}} p(x, \theta|y_{0:k})
\]

1. **Smoothing Step**

\[
y_{f,(m)}^k = H_k \left( \mathcal{M}_{k-1}(x_{a,(m)}^{k-1}, \theta^{(m)}_{k-1}) + u_{k-1}^{(m)} \right) + v_{k}^{(m)} ; \quad v_{k}^{(m)} \sim \mathcal{N}(0, R_k)
\]

\[
x_{k-1}^{s,(m)} = x_{k-1}^{a,(m)} + P_{x_{k-1}^{a,(m)}} y_{k}^{f} P_{y_{k}^{f}}^{-1} \left( y_{k} - y_{f,(m)}^{k} \right)
\]

\[
\theta^{(m)}_{k} = \theta^{(m)}_{k-1} + P_{\theta_{k-1}^{(m)}} y_{k}^{f} P_{y_{k}^{f}}^{-1} \left( y_{k} - y_{f,(m)}^{k} \right)
\]

2. **Analysis Step**

\[
x_{n}^{a,(m)} = \mathcal{M}_{k-1} \left( x_{k-1}^{s,(m)}, \theta^{(m)}_{k} \right) + u_{k-1}^{(m)} ; \quad u_{k-1}^{(m)} \sim \mathcal{N}(0, Q_{k-1})
\]
# Computational Complexity

**Table:** Approximate cost assuming $N_y << N_x$

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time-update</th>
<th>Measurement-update</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint-EnKF</td>
<td>$NN_e (C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
<tr>
<td>Dual-EnKF</td>
<td>$NN_e (2C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
<tr>
<td>Joint-EnKF$_{OSA}$</td>
<td>$NN_e (2C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
</tbody>
</table>
Computational Complexity

Table: Approximate cost assuming $N_y << N_x$

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time-update</th>
<th>Measurement-update</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint-EnKF</td>
<td>$NN_e (C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
<tr>
<td>Dual-EnKF</td>
<td>$NN_e (2C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
<tr>
<td>Joint-EnKF_{OSA}</td>
<td>$NN_e (2C_x + C_\theta)$</td>
<td>$NN_e C_y + NN_e^2 (N_x + N_\theta)$</td>
<td>$2NN_e (S_x + S_\theta)$</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\mathbf{x}^{a,(m)}_k &\quad \text{Dual-EnKF} \\
&\quad = \mathbf{M}_{k-1} \left( \mathbf{x}^{a,(m)}_{k-1}, \theta^{(m)}_k \right) + \mathbf{P}_{x_k^f} \mathbf{H}_k^T \times \mu^{(m)}_k \\
\mathbf{x}^{a,(m)}_k &\quad \text{Joint-EnKF}_{OSA} \\
&\quad = \mathbf{M}_{k-1} \left( \mathbf{x}^{a,(m)}_{k-1} + \mathbf{P}_{x_{k-1}^a} \mathbf{y}_k \times \nu^{(m)}_k, \theta^{(m)}_k \right) \\
&\quad \quad \quad \quad \quad \quad + \mathbf{M}_{k-1} \left( \mathbf{x}^{s,(m)}_{k-1}, \theta^{(m)}_k \right)
\end{align*}
\]
Testing with 1D Ecosystem Model (NPZ)

Experimental setup

- Cycles of phytoplankton blooms in a water column (Eknes and Evensen 2002)
- 4-Years simulation period, 20 layers
- Layer depth: 10m, Time step: 1day

DA framework

- (Stochastic) EnKF, 80 members
- Twin experiments
- **State variables**: Nutrients ($N$), Phytoplankton ($P$), Zooplankton ($H$)
- **Parameters**: Metabolic Loss Rate ($r$), Grazing Efficiency ($f$), Loss to Carnivores ($g$)

**Fig**: Reference run solution
System Configuration and Scenarios

Initialization

- Reference run is initialized from the output of a spin-up solution (5 years)
- The parameters are log-normally distributed in space around specified original values with 50% error
- The state members are assumed to follow a Gaussian distribution

Observations

- Observe the concentration of $N$, $P$, and $H$ every 5 days
- 3 different observation networks: from all layers (20), half (10), and quarter (5)
- Observational error: $\epsilon_k \sim \mathcal{N}(0, \sigma = 0.3 \times y_k)$

Assimilation scenarios

- 4-Years assimilation period
- Experiments repeated 20 times for robustness
- Diagnostics (RMS, ...) averaged over the experiments
**State Estimates: Time-evolution RMS**

- RMS errors for the nutrients are comparable
- Most improvements of the proposed Joint-EnKF-OSA are given by the estimates of Phytoplanktons and Zooplanktons
- The standard joint and dual schemes behave poorly during the spring bloom
- Similar behavior is observed when assimilating half and quarter of the observations

**Figure:** Time-evolution of RMS; observing all layers.
## State Estimates: Average RMS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.0753</td>
<td>0.0722</td>
<td>0.0614</td>
<td>18%</td>
<td>15%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0889</td>
<td>0.1011</td>
<td>0.0765</td>
<td>14%</td>
<td>24%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.1184</td>
<td>0.1207</td>
<td>0.0893</td>
<td>25%</td>
<td>26%</td>
</tr>
</tbody>
</table>
## State Estimates: Average RMS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.0753</td>
<td>0.0722</td>
<td>0.0614</td>
<td>18%</td>
<td>15%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0889</td>
<td>0.1011</td>
<td>0.0765</td>
<td>14%</td>
<td>24%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.1184</td>
<td>0.1207</td>
<td>0.0893</td>
<td>25%</td>
<td>26%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.0498</td>
<td>0.0517</td>
<td>0.0282</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0578</td>
<td>0.0604</td>
<td>0.0332</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.0658</td>
<td>0.0643</td>
<td>0.0381</td>
<td>42%</td>
<td>41%</td>
</tr>
</tbody>
</table>
# State Estimates: Average RMS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.0753</td>
<td>0.0722</td>
<td>0.0614</td>
<td>18%</td>
<td>15%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0889</td>
<td>0.1011</td>
<td>0.0765</td>
<td>14%</td>
<td>24%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.1184</td>
<td>0.1207</td>
<td>0.0893</td>
<td>25%</td>
<td>26%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.0498</td>
<td>0.0517</td>
<td>0.0282</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0578</td>
<td>0.0604</td>
<td>0.0332</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.0658</td>
<td>0.0643</td>
<td>0.0381</td>
<td>42%</td>
<td>41%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Joint-EnKF</th>
<th>Dual-EnKF</th>
<th>EnKF-OSA</th>
<th>Imp. JE</th>
<th>Imp. DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.0299</td>
<td>0.0307</td>
<td>0.0135</td>
<td>55%</td>
<td>56%</td>
</tr>
<tr>
<td>Half</td>
<td>0.0347</td>
<td>0.0363</td>
<td>0.0162</td>
<td>53%</td>
<td>55%</td>
</tr>
<tr>
<td>Quarter</td>
<td>0.0378</td>
<td>0.0375</td>
<td>0.0180</td>
<td>52%</td>
<td>52%</td>
</tr>
</tbody>
</table>
The proposed scheme suggests smaller ensemble spreads; larger confidence in the resulting estimates.

Unlike the standard schemes, less over-shooting is observed at the bloom time.
The proposed scheme suggests smaller ensemble spreads; larger confidence in the resulting estimates.

Unlike the standard schemes, less over-shooting is observed at the bloom time.

Better maintaining of the ensemble spread over time:
Parameter Estimates: Plant metabolic loss (r)

- All layers are observed
- Quick convergence towards the target value
- No significant difference between the schemes
Parameter Estimates: Grazing efficiency (f)

- 10 layers are observed
- First bloom: Joint and Dual-EnKFs impose large corrections in opposite direction
- Significant improvement is obtained using the proposed Joint-EnKF_{OSA} scheme
Parameter Estimates: Loss to carnivores (g)

- 5 layers are observed
- The Dual-EnKF performs better than the Joint-EnKF
- Bloom times: Joint and Dual-EnKFs impose corrections in different directions
- The proposed Joint-EnKF$_{OSA}$ scheme is the most accurate with quick convergence
Parameter Estimates: All assimilation runs

- 5 layers are observed
- 20 runs: The proposed Joint-EnKF_{OSA} scheme is robust and much more accurate than the other schemes
Impact of truncation on the estimation

- 5 layers are observed, one assimilation run
- High truncation observed using the Joint and the Dual-EnKFs: Depletion of the herbivores ensemble; experience large correction on parameters in wrong directions
- The proposed scheme shows less truncation thanks to its dynamically more consistent updating algorithm
Concluding Remarks

• Data assimilation in ocean ecosystem models is challenging given its highly nonlinear character and the poorly known parameters

• Standard assimilation techniques might become inconsistent under complex scenarios

• We propose a smoothing-based joint ensemble Kalman filter in which the measurement and the time update steps are reversed
  ▶ More accurate state and parameter estimates
  ▶ More robust to assimilation scenarios: less truncation of “unphysical” ensemble variables

• Currently being employed in the atlantic system assimilating real physical and biological data

• Future research: work with different ensemble sizes for the state and parameters!