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

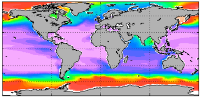
June, 2015

Context: Ocean Ecosystem Modeling

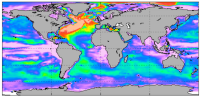
Coupled models

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

Surface silicate



Anthropogenic Carbon

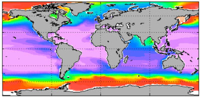


Context: Ocean Ecosystem Modeling

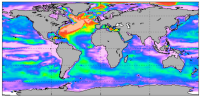
Coupled models

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

Surface silicate



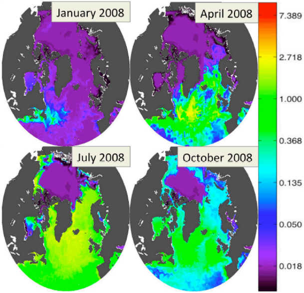
Anthropogenic Carbon



Biological data

- Satellite: surface chlorophyll-a
- In-situ: Nutrients concentrations, pCO_2, ..

CHLa (mg/m^3)

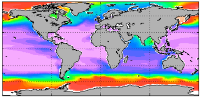


Context: Ocean Ecosystem Modeling

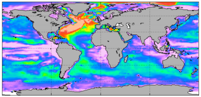
Coupled models

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

Surface silicate



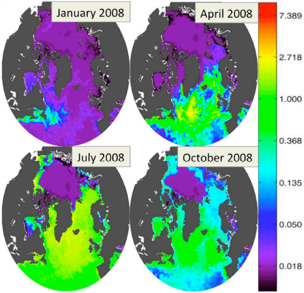
Anthropogenic Carbon



Biological data

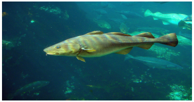
- Satellite: surface chlorophyll-a
- In-situ: Nutrients concentrations, pCO_2, ..

CHLa (mg/m^3)



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 $\Rightarrow$ Need to find $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right)$ recursively in time using an EnKF

Challenges and Motivation

Objective $\Rightarrow$ Need to find $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right)$ 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 $\Rightarrow$ Need to find $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right)$ 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 $\Rightarrow$ Need to find $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k})$ 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 $\Rightarrow$ Need to find $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k})$ 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 $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \to p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\ \mathcal{F}\ } \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\ \mathcal{A}\ } \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k}\right) \end{array}}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_k | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right) \end{array}}$$

- 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 $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\mathbf{z}_k | \mathbf{y}_{0:k})$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k}) \end{array}}$$

- 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 $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\theta_k | \mathbf{y}_{0:k}) \cdot p(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k})$.
$\rightsquigarrow$ *Update the parameters before the state:*

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\leadsto$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k}\right) \end{array}}$$

- 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 $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\theta_k | \mathbf{y}_{0:k}\right) \cdot p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k}\right)$.
$\leadsto$ *Update the parameters before the state:*

$$\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\mathbf{z}_k | \mathbf{y}_{0:k})$.
$\leadsto$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k}) \end{array}}$$

- 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 $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\theta_k | \mathbf{y}_{0:k}) \cdot p(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k})$.
$\leadsto$ *Update the parameters before the state:*

$$\boxed{p(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1})} \xrightarrow{\mathcal{A}^{\theta}} \boxed{p(\theta_k | \mathbf{y}_{0:k})}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k}\right) \end{array}}$$

- 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 $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \rightarrow p\left(\theta_k | \mathbf{y}_{0:k}\right) \cdot p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update the parameters before the state:*

$$\begin{tikzcd} \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p\left(\theta_k | \mathbf{y}_{0:k}\right)} \end{tikzcd}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \to p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k}\right) \end{array}}$$

- 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 $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \to p\left(\theta_k | \mathbf{y}_{0:k}\right) \cdot p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update the parameters before the state:*

$$\begin{tikzcd} \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^\theta"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p\left(\theta_k | \mathbf{y}_{0:k}\right)} \arrow[r, "\mathcal{F}"] & \boxed{p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k-1}\right)} \end{tikzcd}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\mathbf{z}_k | \mathbf{y}_{0:k})$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}) \\ \equiv p(\mathbf{z}_{k} | \mathbf{y}_{0:k}) \end{array}}$$

- 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 $p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}) \to p(\theta_k | \mathbf{y}_{0:k}) \cdot p(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k})$.
$\rightsquigarrow$ *Update the parameters before the state:*

$$\begin{tikzcd} \boxed{p(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1})} \arrow[r, "\mathcal{A}^{\theta}"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p(\theta_k | \mathbf{y}_{0:k})} \arrow[r, "\mathcal{F}"] & \boxed{p(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k-1})} \arrow[d, "\mathcal{A}^{\mathbf{x}}"] \\ & & \boxed{p(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k})} \end{tikzcd}$$

State-Parameters Estimation (Standard Techniques)

Joint-EnKF: Classical state-space augmented form $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \to p\left(\mathbf{z}_k | \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update both the state and parameters simultaneously:*

$$\boxed{\begin{array}{c} p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k-1} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{F}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k-1}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k-1}\right) \end{array}} \xrightarrow{\mathcal{A}} \boxed{\begin{array}{c} p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right) \\ \equiv p\left(\mathbf{z}_{k} | \mathbf{y}_{0:k}\right) \end{array}}$$

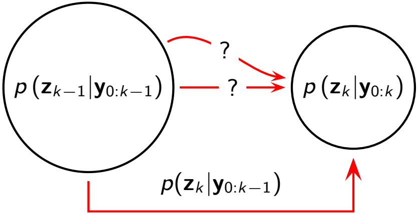
- Could yield to significant inconsistency (Wen and Chen, 2006)
- Might be subject to stability and tractability issues (Moradkhani et al., 2005; Wang et al., 2009)

Dual-EnKF: Separate the densities $p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right) \to p\left(\theta_k | \mathbf{y}_{0:k}\right) \cdot p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k}\right)$.
$\rightsquigarrow$ *Update the parameters before the state:*

$$\begin{tikzcd} \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p\left(\theta_k | \mathbf{y}_{0:k}\right)} \arrow[r, "\mathcal{F}"] & \boxed{p\left(\mathbf{x}_k | \theta_k, \mathbf{y}_{0:k-1}\right)} \arrow[d, "\mathcal{A}^{\mathbf{x}}"] \\ & & \boxed{p\left(\mathbf{x}_k, \theta_k | \mathbf{y}_{0:k}\right)} \end{tikzcd}$$

- <u>More expensive:</u> requires 2 forward model integrations!

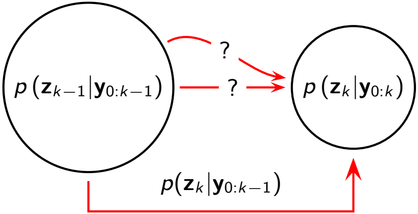
One-Step-Ahead Smoothing-based Joint-EnKF



Alternative formulation:

- The classical path that involves the forecast pdf $p(\mathbf{z}_k|\mathbf{y}_{0:k-1})$ when moving from the analysis pdf $p(\mathbf{z}_{k-1}|\mathbf{y}_{0:k-1})$ to the analysis pdf at the next time $p(\mathbf{z}_k|\mathbf{y}_{0:k})$ is not unique!

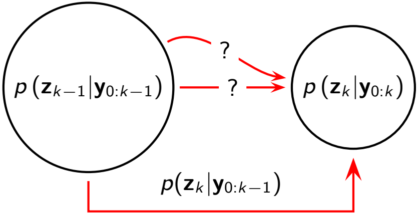
One-Step-Ahead Smoothing-based Joint-EnKF



Alternative formulation:

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

One-Step-Ahead Smoothing-based Joint-EnKF



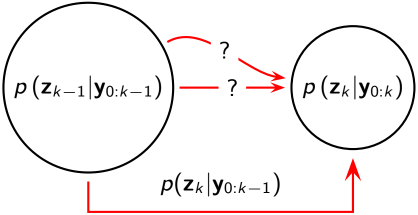
Alternative formulation:

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

1. Smoothing Step: $p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k})$ is first computed using likelihood $p(\mathbf{y}_k|\mathbf{x}_{k-1},\theta)$:

$$p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k}) \propto p(\mathbf{y}_k|\mathbf{x}_{k-1},\theta)p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k-1})$$

One-Step-Ahead Smoothing-based Joint-EnKF



Alternative formulation:

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

1. Smoothing Step: $p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k})$ is first computed using likelihood $p(\mathbf{y}_k|\mathbf{x}_{k-1},\theta)$:

$$p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k}) \propto p(\mathbf{y}_k|\mathbf{x}_{k-1},\theta)p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k-1})$$

2. Analysis Step: $p(\mathbf{x}_k|\mathbf{y}_{0:k})$ is computed using posteriori transition $p(\mathbf{x}_k|\mathbf{x}_{k-1},\theta,\mathbf{y}_k)$:

$$p(\mathbf{x}_n|\mathbf{y}_{0:k}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1},\theta,\mathbf{y}_k)p(\mathbf{x}_{k-1},\theta|\mathbf{y}_{0:k})d\mathbf{x}_{k-1}d\theta$$

Ensemble Implementation: Joint-EnKF$_{\mathrm{OSA}}$

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

Ensemble Implementation: Joint-EnKF$_{\text{OSA}}$

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

$$\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} \middle| \mathbf{y}_{0:k-1}\right)}$$

Ensemble Implementation: Joint-EnKF$_{\mathsf{OSA}}$

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

$$\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \xrightarrow{\mathcal{A}^{\theta}, \mathcal{S}^{\mathbf{x}}} \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k} | \mathbf{y}_{0:k}\right)}$$

Ensemble Implementation: Joint-EnKF$_{\text{OSA}}$

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

$$\begin{tikzcd}
\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}, \mathcal{S}^{\mathbf{x}}"] \arrow[r, dashed, bend right, "\mathcal{F}"'] & \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k} | \mathbf{y}_{0:k}\right)}
\end{tikzcd}$$

Ensemble Implementation: Joint-EnKF$_{\mathsf{OSA}}$

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

$$
\begin{tikzcd}
\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}, \mathcal{S}^{\mathbf{x}}"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k} | \mathbf{y}_{0:k}\right)} \arrow[r, "\mathcal{F}"] & \boxed{p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right)}
\end{tikzcd}
$$

Ensemble Implementation: Joint-EnKF$_{\mathsf{OSA}}$

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

$$
\begin{tikzcd}
\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}, \mathcal{S}^{\mathbf{x}}"] \arrow[r, dashed, bend right, "\mathcal{F}"'] & \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k} | \mathbf{y}_{0:k}\right)} \arrow[r, "\mathcal{F}"] & \boxed{p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right)}
\end{tikzcd}
$$

1. Smoothing Step

$$\mathbf{y}_k^{f,(m)} = \mathbf{H}_k \left(\mathcal{M}_{k-1}(\mathbf{x}_{k-1}^{a,(m)}, \theta_{|k-1}^{(m)}) + \mathbf{u}_{k-1}^{(m)} \right) + \mathbf{v}_k^{(m)} \quad ; \qquad \mathbf{v}_k^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$$

$$\mathbf{x}_{k-1}^{s,(m)} = \mathbf{x}_{k-1}^{a,(m)} + \mathbf{P}_{\mathbf{x}_{k-1}^{a}, \mathbf{y}_k^f} \mathbf{P}_{\mathbf{y}_k^f}^{-1} \left(\mathbf{y}_k - \mathbf{y}_k^{f,(m)} \right)$$

$$\theta_{|k}^{(m)} = \theta_{|k-1}^{(m)} + \mathbf{P}_{\theta_{|k-1}, \mathbf{y}_k^f} \mathbf{P}_{\mathbf{y}_k^f}^{-1} \left(\mathbf{y}_k - \mathbf{y}_k^{f,(m)} \right)$$

Ensemble Implementation: Joint-EnKF$_{\text{OSA}}$

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

$$\begin{tikzcd}
\boxed{p\left(\mathbf{x}_{k-1}, \theta_{k-1} | \mathbf{y}_{0:k-1}\right)} \arrow[r, "\mathcal{A}^{\theta}, \mathcal{S}^{\mathbf{x}}"] \arrow[r, bend right, dashed, "\mathcal{F}"'] & \boxed{p\left(\mathbf{x}_{k-1}, \theta_{k} | \mathbf{y}_{0:k}\right)} \arrow[r, "\mathcal{F}"] & \boxed{p\left(\mathbf{x}_{k}, \theta_{k} | \mathbf{y}_{0:k}\right)}
\end{tikzcd}$$

1. Smoothing Step

$$\mathbf{y}_k^{f,(m)} = \mathbf{H}_k \left(\mathcal{M}_{k-1}(\mathbf{x}_{k-1}^{a,(m)}, \theta_{|k-1}^{(m)}) + \mathbf{u}_{k-1}^{(m)} \right) + \mathbf{v}_k^{(m)} \quad ; \quad \mathbf{v}_k^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$$

$$\mathbf{x}_{k-1}^{s,(m)} = \mathbf{x}_{k-1}^{a,(m)} + \mathbf{P}_{\mathbf{x}_{k-1}^{a}, \mathbf{y}_k^f} \mathbf{P}_{\mathbf{y}_k^f}^{-1} \left(\mathbf{y}_k - \mathbf{y}_k^{f,(m)} \right)$$

$$\theta_{|k}^{(m)} = \theta_{|k-1}^{(m)} + \mathbf{P}_{\theta_{|k-1}, \mathbf{y}_k^f} \mathbf{P}_{\mathbf{y}_k^f}^{-1} \left(\mathbf{y}_k - \mathbf{y}_k^{f,(m)} \right)$$

2. Analysis Step

$$\mathbf{x}_n^{a,(m)} = \mathcal{M}_{k-1} \left(\mathbf{x}_{k-1}^{s,(m)}, \theta_{|k}^{(m)} \right) + \mathbf{u}_{k-1}^{(m)} \quad ; \quad \mathbf{u}_{k-1}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$$

Computational Complexity

Table: Approximate cost assuming $N_y << N_x$

Algorithm	Time-update	Measurement-update	Storage
Joint-EnKF	$NN_e(\mathcal{C}_x + \mathcal{C}_\theta)$	$NN_e\mathcal{C}_y + NN_e^2(N_x + N_\theta)$	$2NN_e(\mathcal{S}_x + \mathcal{S}_\theta)$
Dual-EnKF	$NN_e(2\mathcal{C}_x + \mathcal{C}_\theta)$	$NN_e\mathcal{C}_y + NN_e^2(N_x + N_\theta)$	$2NN_e(\mathcal{S}_x + \mathcal{S}_\theta)$
Joint-EnKF$_{OSA}$	$NN_e(2\mathcal{C}_x + \mathcal{C}_\theta)$	$NN_e\mathcal{C}_y + NN_e^2(N_x + N_\theta)$	$2NN_e(\mathcal{S}_x + \mathcal{S}_\theta)$

Computational Complexity

Table: Approximate cost assuming $N_y << N_x$

Algorithm	Time-update	Measurement-update	Storage
Joint-EnKF	$NN_e\left(\mathcal{C}_x + \mathcal{C}_\theta\right)$	$NN_e\mathcal{C}_y + NN_e^2\left(N_x + N_\theta\right)$	$2NN_e\left(\mathcal{S}_x + \mathcal{S}_\theta\right)$
Dual-EnKF	$NN_e\left(2\mathcal{C}_x + \mathcal{C}_\theta\right)$	$NN_e\mathcal{C}_y + NN_e^2\left(N_x + N_\theta\right)$	$2NN_e\left(\mathcal{S}_x + \mathcal{S}_\theta\right)$
Joint-EnKF$_{OSA}$	$NN_e\left(2\mathcal{C}_x + \mathcal{C}_\theta\right)$	$NN_e\mathcal{C}_y + NN_e^2\left(N_x + N_\theta\right)$	$2NN_e\left(\mathcal{S}_x + \mathcal{S}_\theta\right)$

$$\mathbf{x}_k^{a,(m)} \overset{\text{Dual}-\text{EnKF}}{=} \mathcal{M}_{k-1}\left(\mathbf{x}_{k-1}^{a,(m)}, \theta_{|k}^{(m)}\right) + \overbrace{\mathbf{P}_{\mathbf{x}_k^f}\mathbf{H}_k^T \times \mu_k^{(m)}}^{\text{correction term}}$$

$$\mathbf{x}_k^{a,(m)} \overset{\text{Joint}-\text{EnKF}_{\text{OSA}}}{=} \mathcal{M}_{k-1}\left(\underbrace{\mathbf{x}_{k-1}^{a,(m)} + \overbrace{\mathbf{P}_{\mathbf{x}_{k-1}^a,\mathbf{y}_k^f} \times \nu_k^{(m)}}^{\text{correction term}}}_{\mathbf{x}_{k-1}^{s,(m)}},\ \theta_{|k}^{(m)}\right)$$

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)

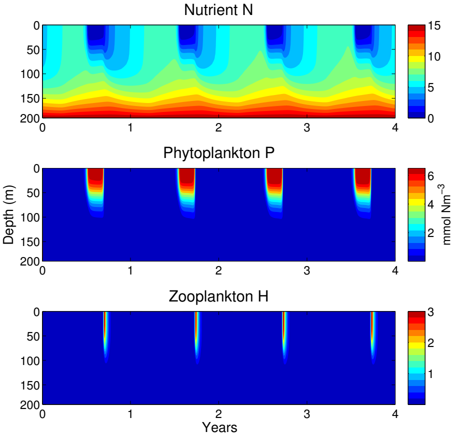


Fig: Reference run solution

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: $\epsilon_k \sim \mathcal{N}\left(0, \sigma = 0.3 \times \mathbf{y}_k\right)$

Assimilation scenarios

- 4-Years assimilation period
- Experiments repeated 20 times for robustness
- Diagnostics (RMS, ...) averaged over the experiments

State Estimates: Time-evolution RMS

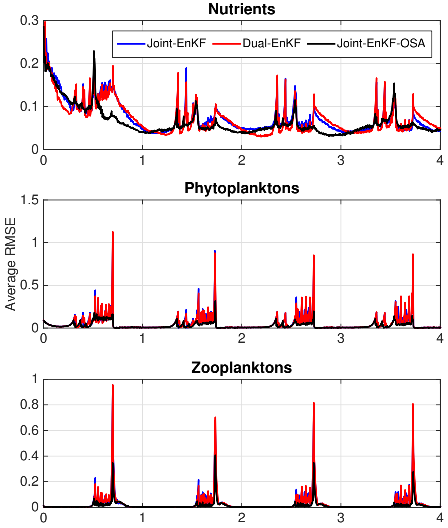


Figure: Time-evolution of RMS; observing all layers.

- RMS errors for the nutrients are comparable
- Most improvements of the proposed Joint-EnKF$_{\text{OSA}}$ 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

State Estimates: Average RMS

(N):

Scenario	**Joint-EnKF**	**Dual-EnKF**	**EnKF-OSA**	**Imp. JE**	**Imp. DE**
All	0.0753	0.0722	**0.0614**	**18%**	**15%**
Half	0.0889	0.1011	**0.0765**	**14%**	**24%**
Quarter	0.1184	0.1207	**0.0893**	**25%**	**26%**

State Estimates: Average RMS

(N):

Scenario	**Joint-EnKF**	**Dual-EnKF**	**EnKF-OSA**	**Imp. JE**	**Imp. DE**
All	0.0753	0.0722	**0.0614**	**18%**	**15%**
Half	0.0889	0.1011	**0.0765**	**14%**	**24%**
Quarter	0.1184	0.1207	**0.0893**	**25%**	**26%**

(P):

Scenario	**Joint-EnKF**	**Dual-EnKF**	**EnKF-OSA**	**Imp. JE**	**Imp. DE**
All	0.0498	0.0517	**0.0282**	**43%**	**45%**
Half	0.0578	0.0604	**0.0332**	**43%**	**45%**
Quarter	0.0658	0.0643	**0.0381**	**42%**	**41%**

State Estimates: Average RMS

(N):

Scenario	Joint-EnKF	Dual-EnKF	EnKF-OSA	Imp. JE	Imp. DE
All	0.0753	0.0722	**0.0614**	**18%**	**15%**
Half	0.0889	0.1011	**0.0765**	**14%**	**24%**
Quarter	0.1184	0.1207	**0.0893**	**25%**	**26%**

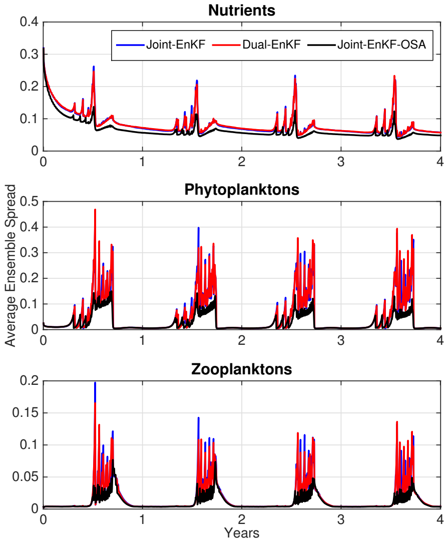
(P):

Scenario	Joint-EnKF	Dual-EnKF	EnKF-OSA	Imp. JE	Imp. DE
All	0.0498	0.0517	**0.0282**	**43%**	**45%**
Half	0.0578	0.0604	**0.0332**	**43%**	**45%**
Quarter	0.0658	0.0643	**0.0381**	**42%**	**41%**

(H):

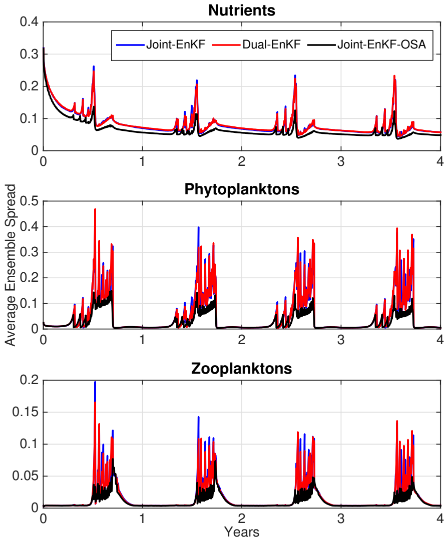
Scenario	Joint-EnKF	Dual-EnKF	EnKF-OSA	Imp. JE	Imp. DE
All	0.0299	0.0307	**0.0135**	**55%**	**56%**
Half	0.0347	0.0363	**0.0162**	**53%**	**55%**
Quarter	0.0378	0.0375	**0.0180**	**52%**	**52%**

State Estimates: Spread

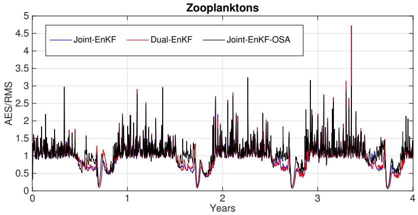


- The proposed scheme suggests smaller ensemble spreads; larger confidence in the resulting estimates
- Unlike the standard schemes, less over-shooting is observed at the bloom time

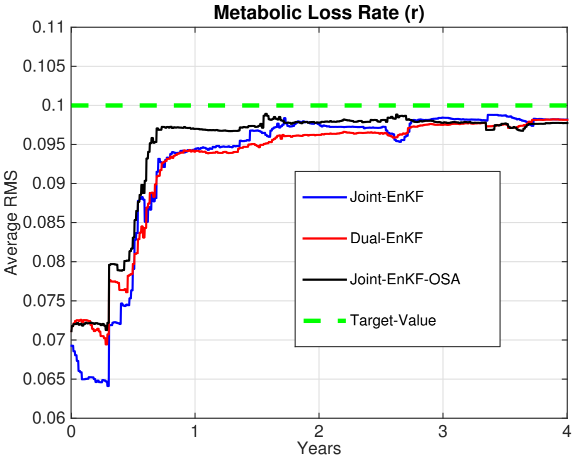
State Estimates: Spread



- 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:

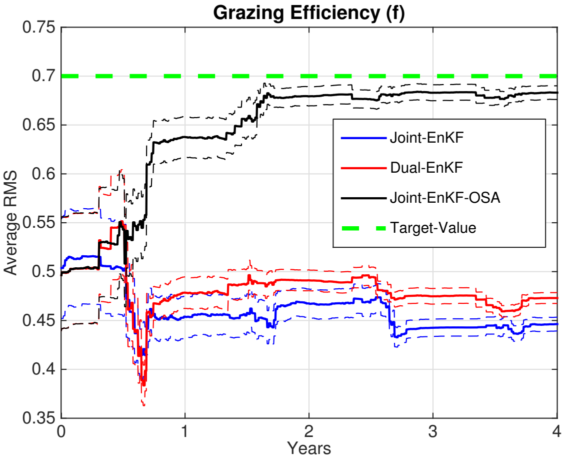


Parameter Estimates: Plant metabolic loss (r)



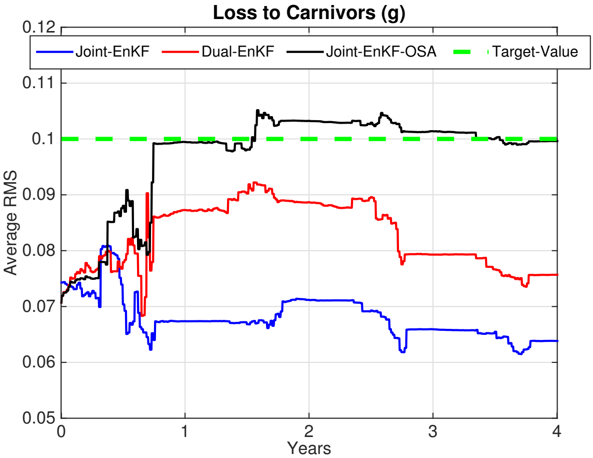
- All layers are observed
- Quick convergence towards the target value
- No significant difference between the schemes

Parameter Estimates: Grazing efficiency (f)



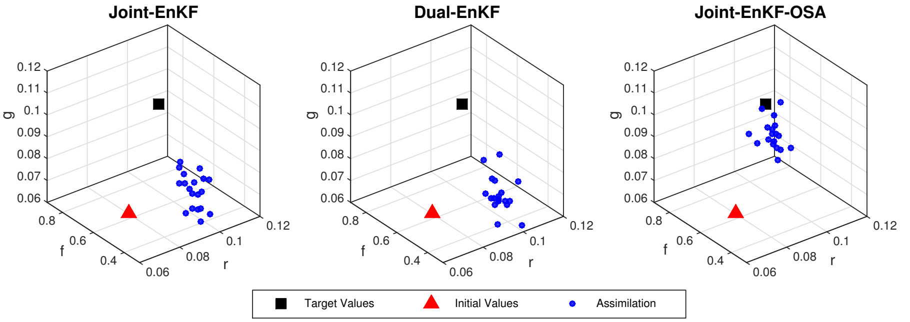
- 10 layers are observed
- First bloom: Joint and Dual-EnKFs impose large corrections in opposite direction
- Significant improvement is obtained using the proposed Joint-EnKF$_{\text{OSA}}$ scheme

Parameter Estimates: Loss to carnivores (g)



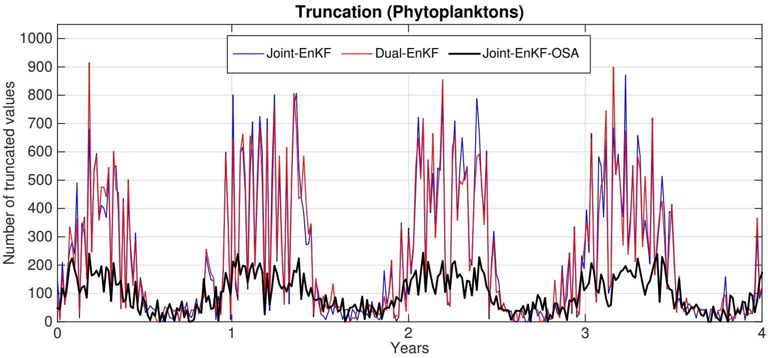
- 5 layers are observed
- The Dual-EnKF performs better than the Joint-EnKF
- Bloom times: Joint and Dual-EnKFs impose corrections in different directions
- The proposed Joint-EnKF$_{\text{OSA}}$ scheme is the most accurate with quick convergence

Parameter Estimates: All assimilation runs



- 5 layers are observed
- 20 runs: The proposed Joint-EnKF$_{\mathrm{OSA}}$ 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!