Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters: application to an ocean ecosystem model

Ehouarn Simon[1], Annette Samuelsen[1], Laurent Bertino[1]
Dany Dumont[2]

[1]Nansen Environmental and Remote Sensing Center, Norway
[2]Institut des sciences de la mer, Université du Québec à Rimouski, Canada

16 November 2012

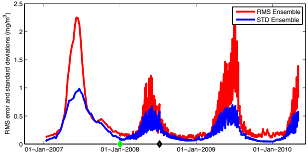
Outline

1. Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters

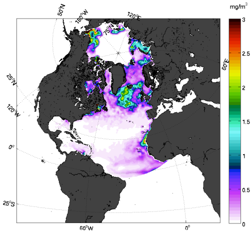
2. Numerical results in GOTM-NORWECOM

Assimilation of ocean color data

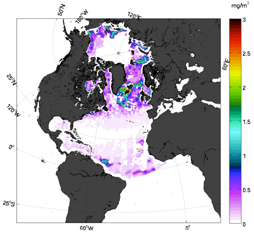
Time evolution of the errors



CHLA: forecast mean (13-08-2008)



CHLA: observations (13-08-2008)



Difficulties

- Nonlinear dynamics and positive variables (tracer concentration).
- Numerous poorly known parameters.
- Seasonal lack of observations (Arctic) and large error (satellite), few in-situ data.

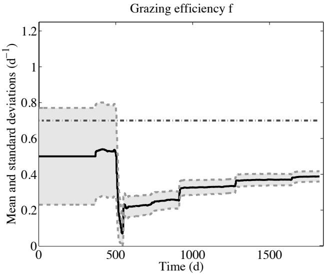
Diet of the herbivorous / predators ?

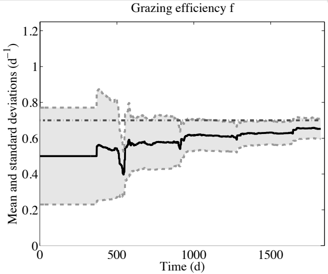
- Zooplankton grazing preferences (relative diet).
  - ▷ No prior knowledge, one set of values for the whole domain.
- Adaptation of zooplankton to their local environment?

Non-Gaussianity and parameter estimation with the EnKF

Truncations of negative values

- Potential depletion of components of the ensemble
  - ▷ Large corrections on parameters in erroneous directions.





Gaussian anamorphosis extension of ensemble-based Kalman filters

- Analysis with transformed variables and observations (Bertino et al., 2003)
- Applicability of the methods in large systems
- Effectiveness of the Gaussian anamorphosis extension of the EnKF to estimate parameters in nonlinear frameworks of positive variables.

Outline

1. Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters

2. Numerical results in GOTM-NORWECOM

Positive sum-to-one constrained parameters $(\pi_i)_{i\in\mathbb{N}_N}$

Estimation

- Constraints: $\forall i = 1:N, \pi_i \geqslant 0$ and $\sum_{i=1}^{N} \pi_i = 1$
  - ▷ Cannot be handled by the EnKF in that form.

Dirichlet distribution of order N

$$\forall i = 1:N, \pi_i = \frac{\phi_i}{\sum_{k=1}^{N} \phi_k} \quad \text{with} \quad \phi_i \sim \Gamma(\theta_i, 1)$$

Gelman's formulation (1995)

$$\forall i = 1:N, \pi_i = \frac{e^{\phi_i}}{\sum_{k=1}^{N} e^{\phi_k}} \quad \text{with} \quad \phi_i \sim \mathcal{N}(\theta_i, \Sigma_i)$$

Positive sum-to-one constrained parameters $(\pi_i)_{i\in\mathbb{N}_N}$

Hyperspherical coordinates: $N-1$ parameters

$$\begin{cases} \pi_1 = \cos^2(\frac{\pi}{2}\phi_1) \\ \forall i = 2:N-1, \\ \pi_i = \prod_{k=1}^{i-1} \sin^2(\frac{\pi}{2}\phi_k)\cos^2(\frac{\pi}{2}\phi_i) \\ \pi_N = \prod_{k=1}^{N-2} \sin^2(\frac{\pi}{2}\phi_k)\sin^2(\frac{\pi}{2}\phi_{N-1}) \end{cases}$$

- with $(\phi_i)_{i=1=N-1}$ distributed on the segment line $[0,1]$

Inversion of the hyperspherical coordinate system

- Recursively.
- Information on the distribution of the $(\pi_i)_{i=1:N}$ or available samples:
  - ▷ Prior values for the $(\phi_i)_{i=1:N-1}$.

Expected values and variances of the $(\pi_i)_{i=1:N}$

Tuning of the parameters of the distributions of the $(\phi_i)_{i=1:N-1}$

- $\forall i = 1 : N-1, \quad E[\pi_i] = m_i, \quad \text{var}(\pi_i) = \sigma_i^2$
- Assumption: $(\phi_i)_{i=1:N-1} \sim (\mathcal{D}_i([0,1], \Theta_i))_{i=1:N-1}$ are independent.

Expected values of the $(\pi_i)_{i=1:N-1}$

$$\begin{cases} \frac{1}{4}(E[e^{j\pi\phi_1}] + E[e^{-j\pi\phi_1}]) = m_1 - \frac{1}{2} \\ \forall i = 2 : N-1, \frac{1}{4}(E[e^{j\pi\phi_i}] + E[e^{-j\pi\phi_i}]) = \frac{m_i}{1 - \sum_{k=1}^{i-1} m_k} - \frac{1}{2} \end{cases}$$

Variances of the $(\pi_i)_{i=1:N-1}$

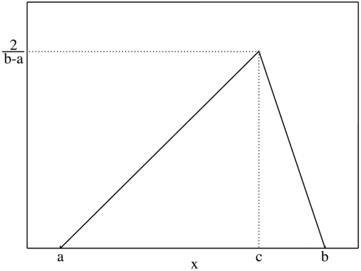
$$\begin{cases} \frac{1}{16}(E[e^{2j\pi\phi_1}] + E[e^{-2j\pi\phi_1}]) = -\frac{3}{8} + \sigma_1^2 + m_1^2 - \frac{1}{4}(E[e^{j\pi\phi_1}] + E[e^{-j\pi\phi_1}]) \\ \forall i = 2 : N-1, \frac{1}{16}(E[e^{2j\pi\phi_i}] + E[e^{-2j\pi\phi_i}]) = -\frac{3}{8} - \frac{1}{4}(E[e^{j\pi\phi_i}] + E[e^{-j\pi\phi_i}]) \\ \qquad + \frac{\sigma_i^2 + m_i^2}{\sum_{k=1}^{i} (-2)^{k-1}(\sigma_{i-k}^2 + m_{i-k}^2) \prod_{l=1}^{k-1} \frac{1}{4}(E[e^{j\pi\phi_{i-l}}] + E[e^{-j\pi\phi_{i-l}}])} \end{cases}$$

Example with the triangular distribution

Triangular distribution: $\phi_i \sim \mathcal{T}(0, 1, c_i)$

- Tuning of the mode c_i.
- Characteristic function:

$$\forall t \in \mathbb{R}, \quad E[e^{jt\phi_i}] = -2\frac{(1-c_i) - e^{jc_it} + c_ie^{jt}}{\pi^2 c_i(1-c_i)}$$



Equal preferences: $E[\pi_i] = \frac{1}{N}$

- A system of $N-1$ nonlinear equations:

$$\forall i = 1 : N-1, \quad \frac{\cos(\pi c_i) + 2c_i - 1}{\pi^2 c_i(1-c_i)} + \frac{N-i-1}{2(N-i+1)} = 0.$$

- $N > 3$: non existence of solutions.

Outline

1. Estimation of positive sum-to-one constrained parameters with ensemble-based Kalman filters

2. Numerical results in GOTM-NORWECOM

GOTM-NORWECOM-ASSIM

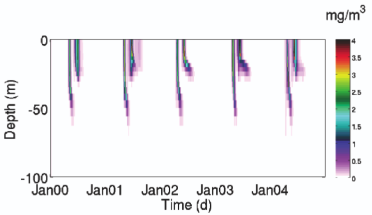
Assimilation

- Deterministic ensemble Kalman filter (DEnKF, Sakov and Oke, 2008).
- State vector: biogeochemical state vector and parameters.
- 100 members.
- Gaussian anamorphosis (Bertino et al., 2003) for the biogeochemical state variables (and parameters for the spherical formulation) .

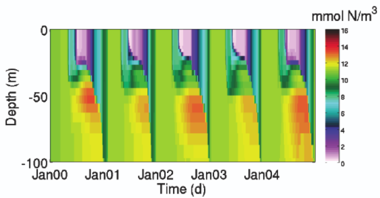
Reference solution $\mathbf{x}^t$

- Deterministic simulation with different preferences:
  - ▷ Mesozooplankton: $\pi_{dia} = 0.6$, $\pi_{mic} = 0.15$, $\pi_{den} = 0.25$.
  - ▷ Microzooplankton: $\pi_{fla} = 0.6$, $\pi_{den} = 0.15$, $\pi_{dia} = 0.25$.

Chlorophyll



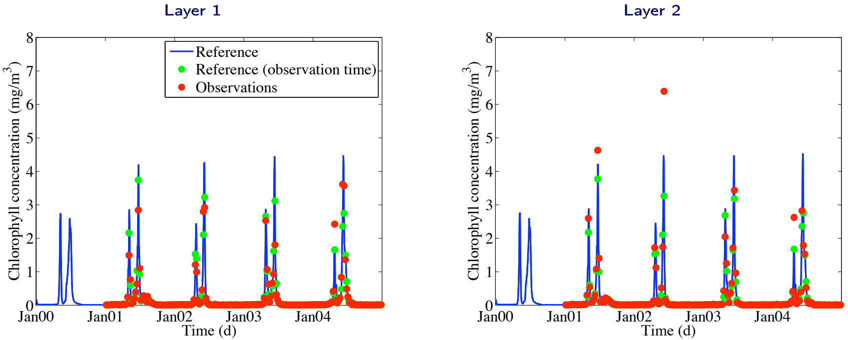
Nitrate



The data assimilation system

Configuration of the experiments

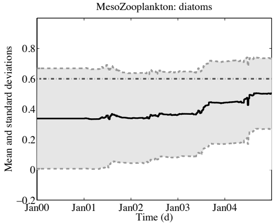
- One year to warm up the ensemble, then four years with assimilation.
- Observations: $\mathbf{y} = \mathbf{x}_P^t * G$, with $G \sim \Gamma(\frac{1}{\sigma_o^2}, \sigma_o^2)$, $\sigma_o = 0.3$.
  - ▷ Surface chlorophyll (2 first layers) every seven days.
- Truncated-Gaussian perturbations in the whole water column every twelve hours: phytoplankton and zooplankton only.
- Robustness of the estimation: experiments repeated 20 times.



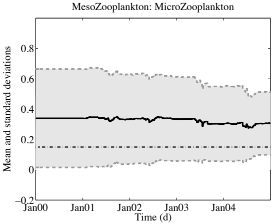
Preferences: Gelman's formulation

$$\pi_i = \frac{e^{\phi_i}}{\sum_{k=1}^{N} e^{\phi_k}}$$

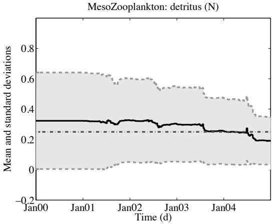
MesoZ π_{dia}



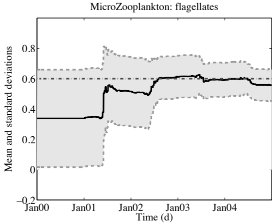
MesoZ π_{mic}



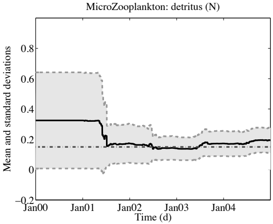
MesoZ π_{den}



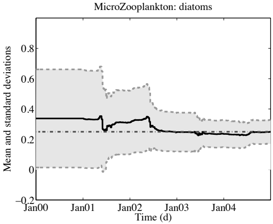
MicroZ π_{fla}



MicroZ π_{den}

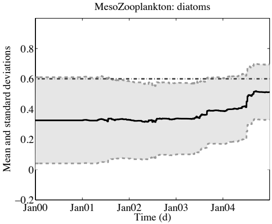


MicroZ π_{dia}

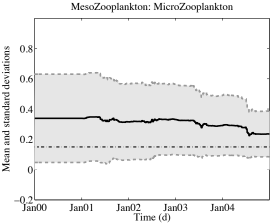


Preferences: spherical formulation

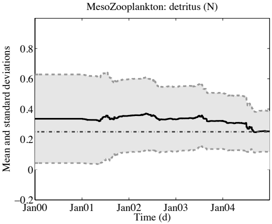
MesoZ π_{dia}



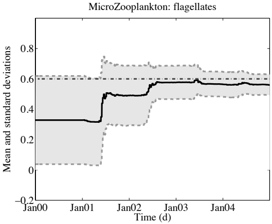
MesoZ π_{mic}



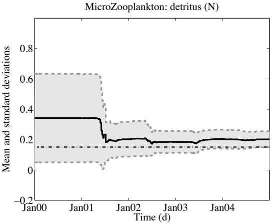
MesoZ π_{den}



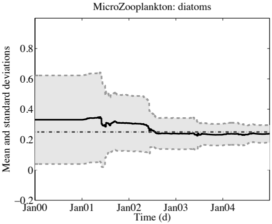
MicroZ π_{fla}



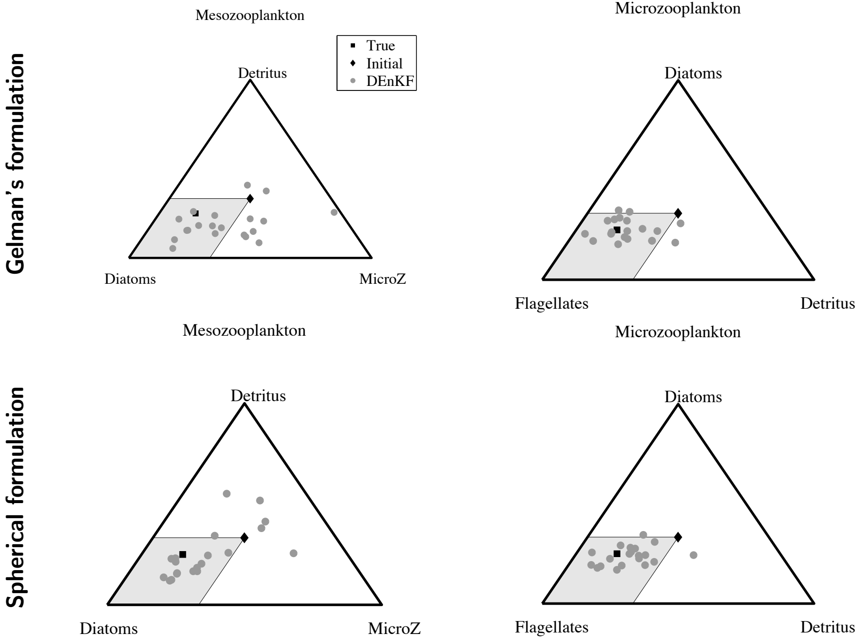
MicroZ π_{den}



MicroZ π_{dia}



Preferences: ternary plots of the final estimates



Spherical formulation: asymmetry of the transformation

Robustness of the results to the matching preferences/transformed parameters

- Same 20 experiments (observation, prior ensembles)
- Microzooplankton: permutation in the assignment preferences/transformed parameters every four experiments.

Mesozooplankton

$$\begin{cases} \pi_{DIA} = \cos^2(\frac{\pi}{2}\phi_1) \\ \pi_{MIC} = \sin^2(\frac{\pi}{2}\phi_1)\cos^2(\frac{\pi}{2}\phi_2) \\ \pi_{DET} = \sin^2(\frac{\pi}{2}\phi_1)\sin^2(\frac{\pi}{2}\phi_2) \end{cases}$$

Microzooplankton

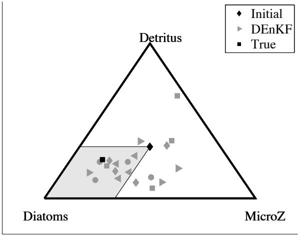
$$\begin{cases} \pi_{DET} = \cos^2(\frac{\pi}{2}\phi_1) \\ \pi_{FLA} = \sin^2(\frac{\pi}{2}\phi_1)\cos^2(\frac{\pi}{2}\phi_2) \\ \pi_{DIA} = \sin^2(\frac{\pi}{2}\phi_1)\sin^2(\frac{\pi}{2}\phi_2) \end{cases}$$

- Microzooplankton: marginal improvement of the estimates of the preferences.
- Mesozooplankton: slight degradation of the estimation.

RMS error (Mesozooplankton.)

Diet	Diatoms	MicroZ	Detritus
Prior (%)	45	120	32
Gelman	35	146.6	44
Spherical	33.3	86.7	48
Spher. + perm.	36.7	133.3	52

Mesozooplankton



Conclusion and perspectives

- Two approaches for estimating positive sum-to-one constrained parameters.
  - ▷ Gelman's formulation: Gaussian parameters, symmetry of the transformation, N transformed parameters.
  - ▷ Spherical formulation: $N-1$ parameters to estimate, asymmetry of the transformation, distribution of the transformed parameters?
- Preliminary experiments indicate that the zooplankton grazing preferences could be estimated with both approaches.
- Quite simple framework.
- Further investigations have to be done.
  - ▷ Distribution.
  - ▷ Additional parameters to estimate.
  - ▷ Real observations (Mike station).

Thank you!

Bibliography

- **Bertino L., Evensen G. and Wackernagel H.:** Sequential Data Assimilation Techniques in Oceanography, *International Statistical Reviews*, 71, 223-241, 2003.
- **Bocquet M., Pires C.A. and Wu L.**: Beyond Gaussian statistical modeling in geophysical data assimilation, *Monthly Weather Review*, 138, 8, 2997-3023, 2010.
- **Doron M., Brasseur P. and Brankart J.-M.**: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model: Twin experiments. *Journal of Marine Systems*, 87 (3-4), 194-207, 2011.
- **Gelman A.**: Method of Moments Using Monte Carlo Simulation. *Journal of Computational and Graphical Statistics*, 4, 1, 36-54, 1995.
- **Gelman A., Bois F. and Jiang J.**: Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions. *Journal of the American Statistical Association*, 91, 436, 1400-1412, 1996.
- **Simon E. and Bertino L**: Gaussian anamorphosis extension of the DEnKF for combined state and parameter estimation: application to a 1D ocean ecosystem model, *Journal of Marine Systems*, 89, 1-18, 2012.
- **Simon E., Samuelsen A., Bertino L. and Dumont D.**: Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF: a twin experiments, *Ocean Science*, 8, 587-602, 2012.

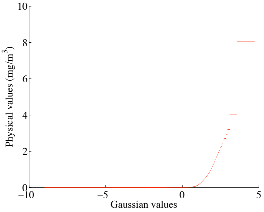
Construction of a monovariate anamorphosis: $Z(x) = \psi(Y(x))$

Empirical anamorphosis

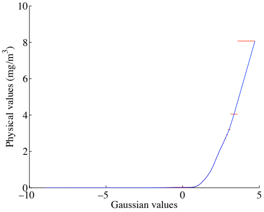
- Empirical marginal distribution of $Z(x)$: sample $(z_i)_{i=1:N}$.
- Cumulative distribution function of $Y(x)$: G.

$$\psi(y) = \sum_{i=1}^{N} z_i 1_{]G^{-1}(\frac{i-1}{N}),G^{-1}(\frac{i}{N})]}(y)$$

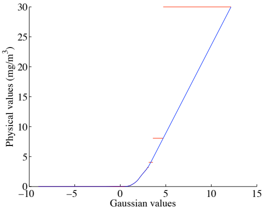
1- Empirical anamorphosis



2- Interpolation



3- Definition of the tails



Major issues

- Choice of the physical data set: specification of the likely and unlikely values.
- Dependance on a reference model run (model bias, extreme events).