

This discussion paper is/has been under review for the journal Ocean Science (OS).
Please refer to the corresponding final paper in OS if available.

Estimation of positive sum-to-one constrained zooplankton grazing preferences with the DEnKF: a twin experiment

E. Simon^{1,2}, A. Samuelsen^{1,2}, L. Bertino^{1,2}, and D. Dumont³

¹Nansen Environmental and Remote Sensing Center, Norway

²Bjerknes Center for Climate Research, Norway

³Institut des sciences de la mer, Université du Québec à Rimouski, Canada

Received: 28 February 2012 – Accepted: 6 March 2012 – Published: 19 March 2012

Correspondence to: E. Simon (ehouarn.simon@nersc.no)

Published by Copernicus Publications on behalf of the European Geosciences Union.

OSD

9, 1085–1121, 2012

Estimation of the
zooplankton grazing
preferences with the
DEnKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

We consider the estimation of the grazing preferences parameters of zooplankton in ocean ecosystem models with ensemble-based Kalman filters. These parameters are introduced to model the relative diet composition of zooplankton that consists of phytoplankton, small size-classes of zooplankton and detritus. They are positive values and their sum is equal to one. However, the sum-to-one constraint cannot be guaranteed by ensemble-based Kalman filters when parameters are bounded. Therefore, a reformulation of the parameterization is proposed. We investigate two types of variables transformations for the estimation of positive sum-to-one constrained parameters that lead to the estimation of new set of parameters with normal or bounded distributions. These transformations are illustrated and discussed with twin experiments performed with the 1-D coupled model GOTM-NORWECOM with Gaussian anamorphosis extensions of the deterministic ensemble Kalman filter (DEnKF).

1 Introduction

The development of numerical ocean biogeochemical models over the last two decades has led to more and more complex representations of the interactions between the different trophic levels, notably between different plankton species at the base of the food chain. While the diet of zooplankton is relatively simply represented in the earliest NPZD models – the unique zooplankton group (Z) is feeding only on the unique phytoplankton group (P), see for example Evans and Parslow (1985) – the addition of multiple plankton functional types (PFT) for the phyto- and zooplankton aiming at representing different plankton species having different functions in the ecosystem (e.g. diatoms, calcifying algae or microzooplankton) leads to more complex diets and grazing preferences must be added. These parameters are always positive, and although not compulsory, usually add up to one. We refer to the review of Gentleman et

OSD

9, 1085–1121, 2012

Estimation of the zooplankton grazing preferences with the DEnKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

al. (2003) for more details concerning the common mathematical formulations of the zooplankton grazing in ocean biological models and their impact on model dynamics.

Grazing preferences specify the direction of the feeding in the space of foods, and so, the direction of the transfer from PFTs (the food) to the zooplankton PFTs (the feeder).

5 Therefore, their impact on the distribution of the different PFTs obtained from a model simulation can be significant. For example, Buitenhuis et al. (2010) observed in their global biogeochemical model that “the phytoplankton functional type distributions and the proportions of primary production that are exported or remineralized” were sensitive to the microzooplankton grazing preferences. In the same way, Buitenhuis et al. (2006) 10 conclude their work by suggesting that the representation of mesozooplankton would notably benefit from the improvement of their grazing preferences by taking into account the food quality. For large scale applications like configurations covering a whole ocean basin, this results in the potential need of a fine spatial tuning of the grazing preferences in order to take into account the adaptation of zooplankton species to their 15 local environments (Gentleman et al., 2003). Direct measurements of grazing preferences for the different zooplankton species would help to optimize the model parameters representing these preferences. However, field data are sparse, the information provided by the experiments realized in laboratory do not cover the large spectrum of conditions found in nature (Buitenhuis et al., 2010) and available observations might 20 not be consistent with each other (Buitenhuis et al., 2006).

Multivariate data assimilation methods like ensemble-based Kalman filters make the estimation of variables and parameters that are not observed possible. State variables and parameters can be estimated simultaneously simply by augmenting the state vector with the parameters to estimate (Anderson, 2001; Evensen, 2009). However, 25 the efficient application of ensemble-based data assimilation methods like the Ensemble Kalman Filter (EnKF; Evensen, 1994, 2003) to ocean ecosystem models is a challenging issue. Beside the nonlinearity of the model, most variables and parameters are strictly positive, producing non-Gaussian state and parameter distributions thereby breaking an important assumption of the linear analysis, and leading to a loss

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

of optimality of Kalman filters. A solution to perform Kalman filter estimation of non-Gaussian variables is the introduction of the Gaussian anamorphosis as suggested by Bertino et al. (2003). This approach has proven to be easily applicable in realistic configurations (Simon and Bertino, 2009) and allows the estimation of biased parameters (Doron et al., 2011; Zhou et al., 2011; Simon and Bertino, 2012).

In this study, we focus on the problem of estimating positive sum-to-one constrained parameters. Our aim is to assess the ability of ensemble-based Kalman filters to estimate zooplankton grazing preferences in ocean biogeochemical models. To overcome the issues that ensemble-based Kalman filters can not guarantee the sum-to-one constraint when a constraint of positiveness applies on the parameters, we investigate two reformulations for which these two constraints are implicit.

The outline of the paper is as follows. We present the different changes of variables for the estimation of positive sum-to-one constrained parameters in Sect. 2. We describe our experimental framework in Sect. 3. Results of the methods are discussed in Sect. 4, and we present our conclusion in Sect. 5.

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

In this section, we describe the general problem of estimating positive sum-to-one constrained parameters with ensemble-based Kalman filters and the issues raised by these constraints. We present a formulation previously suggested by Gelman (1995) to estimate positive sum-to-one constrained parameters in the framework of pharmacokinetics (Gelman et al., 1996). Because the number of food preferences that need to be calibrated can be large in complex ocean biological models (numerous different feeding and fed species), we aim at reducing the number of parameters to estimate. For this reason, we suggest a new formulation that introduces a change of variables based on hyperspherical coordinates.

2.1 Definition of the problem

Let $(\pi_i)_{i=1:N}$ be the N parameters that we wish to estimate. They are positive

$$\forall i = 1:N \quad \pi_i \geq 0, \quad (1)$$

and their sum is equal to one

$$\sum_{i=1}^N \pi_i = 1. \quad (2)$$

They can be estimated with ensemble-based Kalman filter by augmenting the analysis state vector with these parameters. Unfortunately, the conservation of linear properties intrinsic to the ensemble Kalman filter (Evensen, 2003) is not guaranteed for the parameters due to the constraint of positiveness. The truncation of negative values that results from the Kalman analysis can lead to parameter estimates that do not respect the linear sum-to-one property (Eq. 2). Even if the Gaussian anamorphosis extension of ensemble-based Kalman filters makes the estimation possible of positive parameters (Simon and Bertino, 2012), nonlinear transformations do not ensure that they sum to one.

2.2 Dirichlet distribution and Gelman's formulation

A prior distribution for N positive random parameters with the sum-to-one constraint is the Dirichlet distribution of order N . The $(\pi_i)_{i=1:N}$ can be obtained from N independent Gamma distributed random variables $(\phi_i)_{i=1:N}$ as follows:

$$\forall i = 1:N, \pi_i = \frac{\phi_i}{\sum_{k=1}^N \phi_k} \quad \text{with} \quad \phi_i \sim \Gamma(\theta_i, 1). \quad (3)$$

Then, the parameters $(\phi_i)_{i=1:N}$ are estimated by assimilating observation with ensemble-based Kalman filters, and the values of the original parameters $(\pi_i)_{i=1:N}$ are obtained from Eq. (3). Because the parameters $(\phi_i)_{i=1:N}$ are not Gaussian distributed, we suggest to transform them with the Gaussian anamorphosis during the analysis.

Another possibility is to substitute the Gamma distribution for the lognormal distribution as suggested by Gelman (1995).

$$\forall i = 1:N, \pi_i = \frac{e^{\phi_i}}{\sum_{k=1}^N e^{\phi_k}} \quad \text{with} \quad \phi_i \sim \mathcal{N}(\theta_i, \Sigma_i). \quad (4)$$

In that case, the $(\phi_i)_{i=1:N}$ fulfill the Kalman filtering assumption of Gaussian distributed variables and do not require anamorphosis.

Due to the symmetrical roles played by the parameters $(\phi_i)_{i=1:N}$ in both formulations, the estimation of the parameters $(\pi_i)_{i=1:N}$ is not sensitive to the mapping between the parameters $(\phi_i)_{i=1:N}$ and $(\pi_i)_{i=1:N}$ in the change of variables. However, these approaches do not allow for parameters $(\pi_i)_{i=1:N}$ equal to zero, meaning that one food diet would not be completely removed by assimilation. This might be undesirable for large scale configurations for which the feeding can significantly change from a region to another.

2.3 The hyperspherical coordinate system

The $(\pi_i)_{i=1:N}$ can be seen as a position vector in the Cartesian coordinates of a point π in \mathbb{R}^N . Naturally, our idea is to represent this point in another coordinate system. We suggest to introduce $N - 1$ angles $(\phi_i)_{i=1:N-1}$ to represent π in the hyperspherical coordinate system that generalizes the spherical one to the dimension N . An analogy with a coordinates system describing a point on a sphere shows that 2 angles, longitude and latitude, are required to characterize the position of a point at the surface in 3

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



dimensions.

$$\left\{ \begin{array}{l} \pi_1 = \cos^2\left(\frac{\pi}{2}\phi_1\right) \\ \forall i = 2:N - 1, \\ \pi_i = \prod_{k=1}^{i-1} \sin^2\left(\frac{\pi}{2}\phi_k\right) \cos^2\left(\frac{\pi}{2}\phi_i\right) \\ \pi_N = \prod_{k=1}^{N-2} \sin^2\left(\frac{\pi}{2}\phi_k\right) \sin^2\left(\frac{\pi}{2}\phi_{N-1}\right) \end{array} \right. \quad (5)$$

with $(\phi_i)_{i=1:N-1}$ $N - 1$ independent random variables for which the support of the distribution is the segment line $[0, 1]$. By definition, the $(\pi_i)_{i=1:N}$ are positive and it can be easily shown that their sum is equal to one.

One benefit of this approach is the reduction of the number of parameters to estimate from N to $N - 1$. This is certainly worth for complex systems involving numerous unknown parameters, notably if one plans to estimate other parameters besides the $(\pi_i)_{i=1:N}$. However, the estimation of the $(\pi_i)_{i=1:N}$ can be sensitive to the choice of the mapping to the $(\phi_i)_{i=1:N-1}$ because this last ones do not play the same role in the change of variables. For our specific problem of estimating zooplankton grazing preferences, it means that the results might depend on how the different types of food match the $(\pi_i)_{i=1:N}$.

Furthermore, a significant issue lies in the choice of the distribution for the parameters $(\phi_i)_{i=1:N-1}$. We suggest to base this choice on the ability to specify a prior value for the $(\pi_i)_{i=1:N}$ – it means their prior expected value $(E[\pi_i])_{i=1:N}$ – rather than focusing on their distribution. Our main motivation is to be able to start the estimation process with $(\pi_i)_{i=1:N}$ that have the same expected values $\frac{1}{N}$. In our particular framework, this case corresponds to no particular feeding preferences in the diet of the zooplankton species. This leads to the choice of parametric distributions for which the parameters

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



will be tuned according to the prior values of $(\pi_i)_{i=1:N}$ that we want to prescribe. For example, such property discards the use of the uniform distribution $\mathcal{U}(0, 1)$ because $E[\pi_1] = \frac{1}{2}$.

We assume that the parameters $(\phi_i)_{i=1:N-1}$ are independent and follow the same distribution \mathcal{D} involving a set of parameters $(\Theta_i)_{i=1:N-1}$ that may have different values depending on i . The prior values for the expectation of the parameters $(\pi_i)_{i=1:N}$ is obtained by an adequate tuning of the $N - 1$ parameter sets $(\Theta_i)_{i=1:N-1}$. The particular case of a prior with equal expected values for the parameters $(\pi_i)_{i=1:N}$ leads to the resolution of $N - 1$ nonlinear systems:

$\forall i = 1:N - 1$, find Θ_i such that

$$(S_i) \quad \frac{1}{4}(\Phi_{\phi_i}(\pi) + \Phi_{\phi_i}(-\pi)) + \frac{N - i - 1}{2(N - i + 1)} = 0. \quad (6)$$

with Φ_{ϕ_i} the characteristic function of the parameter ϕ_i .

The derivation of these systems is detailed in Appendix A. The existence of solutions to these systems $(S_i)_{i=1:N-1}$ depends on the chosen distribution \mathcal{D} and they are found numerically.

Again, we suggest to transform parameters $(\phi_i)_{i=1:N-1}$ with the anamorphosis functions during the Kalman filter analysis.

Remark: for the particular case of $N = 4$, it is also possible to introduce the Hopf coordinates:

$$\begin{cases} \pi_1 = \cos^2\left(\frac{\pi}{2}\phi_1\right) \sin^2\left(\frac{\pi}{2}\phi_3\right) \\ \pi_2 = \sin^2\left(\frac{\pi}{2}\phi_1\right) \sin^2\left(\frac{\pi}{2}\phi_3\right) \\ \pi_3 = \cos^2\left(\frac{\pi}{2}\phi_2\right) \cos^2\left(\frac{\pi}{2}\phi_3\right) \\ \pi_4 = \sin^2\left(\frac{\pi}{2}\phi_2\right) \cos^2\left(\frac{\pi}{2}\phi_3\right) \end{cases} \quad (7)$$

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



with ϕ_1, ϕ_2, ϕ_3 independent random variables for which the support of the distribution is $[0, 1]$. The simple choice of $\phi_1, \phi_2, \phi_3 \sim \mathcal{U}(0, 1)$ results in equal preferences $(\pi_i)_{i=1:4}$. The generalization to any N parameters is out of the scope of the paper.

3 Experimental framework

3.1 The 1-D ocean ecosystem model

The experiments were performed in a 1-D vertical configuration of the coupled model GOTM-NORWECOM representative of the station Mike (66° N, 2° W) in the North Sea.

The 1-D ocean water column model is the General Ocean Turbulence Model (GOTM; Burchard et al., 1999, 2005; Umlauf and Burchard, 2005) that transports physical quantities with hydrodynamic primitive equations and turbulence schemes. A relaxation towards temperature, salinity and horizontal velocity profiles from the TOPAZ¹ system (Bertino and Lisæter, 2008) is used with a relaxation time of 14 days. The vertical advection velocity is specified to zero. The depth is 2034 m and the model uses a cartesian grid of 55 vertical levels with a minimum thickness of 1 m at the top level, increasing exponentially towards the bottom.

The NORWegian ECOlogical Model system (NORWECOM; Aksnes et al., 1995; Skogen and S iland, 1998) is coupled to GOTM. The current version of this model includes two classes of phytoplankton (diatom and flagellates), two classes of zooplankton (meso- and microzooplankton) derived with the same formulation from the model ECOHAM4 (P atsch et al., 2009), several types of nutrients (inorganic nitrogen, phosphorus and silicon) and detritus (nitrogen, phosphorus), biogenic silica, and oxygen, so that the ecosystem state vector is made of 11 variables. The chlorophyll-a concentration (CHLA) is computed from the model diatoms and flagellates concentrations

¹<http://topaz.nersc.no>

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(DIA and FLA) by the Eq. (8).

$$\text{CHLA} = \frac{\text{DIA} + \text{FLA}}{0.8} \quad (8)$$

The constant conversion factor 0.8 is added to obtain the chlorophyll concentration in mg m^{-3} , the standard unit of data produced from satellite, from the phytoplankton concentration in mmol N m^{-3} . The mesozooplankton (MES) feeds on diatoms (one assumes that the flagellates are too small to be fed on by mesozooplankton), detritus (DEN) and microzooplankton (MIC). The microzooplankton feeds on both classes of phytoplankton (flagellates and diatoms) and on detritus. Both classes of zooplankton have the choice for their food among three variables of the model and compete against each other for feeding on detritus and diatoms. For both classes of zooplankton, the formulation of the grazing $G_{i=1:3}$ on the variable $i = 1:3$ reads:

$$\forall i = 1:3, G_i = g \frac{\pi_i X_i^2}{\sum_{k=1}^3 \pi_k X_k (X_k + K_{1/2})} Z \quad (9)$$

with Z the concentration of meso- or microzooplankton, $(X_k)_{k=1:3}$ the concentration of the different variables they feed on, $(\pi_k)_{k=1:3}$ the grazing preferences, $K_{1/2}$ the half-saturation constant for ingestion by zooplankton and g the zooplankton maximum growth rate. The second order modified Patankar-Runge-Kutta scheme is used for the source and sinks dynamics.

The dynamics of phytoplankton blooms in the first 100 m is illustrated in Fig. 1.

3.2 Data assimilation experiments

In order to assess the performances of the two formulations, twin experiments have been conducted: the true state and the observations are produced by a deterministic

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

simulation of the model involving meso- and microzooplankton grazing preferences that differ from equal preferences. The values of preferences used to build the reference solution will be called “true” values in the following. These values have been arbitrary chosen and are summarized in Table 1.

5 The observations are the chlorophyll in the two first layers of the model and are defined as follows

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n^t \times E, \quad \text{with} \quad E \sim \Gamma\left(\frac{1}{\sigma_o^2}, \sigma_o^2\right), \quad \sigma_o = 0.3 \quad (10)$$

We construct the observations by multiplying the true surface chlorophyll with a Gamma distributed observation error with a standard deviation around 30 % (average should be 1).

10 The state and parameter estimations are conducted jointly by augmenting the state vector with the parameters that are estimated. In this study, the state vector is made of all the vertical components of the ten state variables (the oxygen is not corrected during the analysis) and the parameters $(\phi_i)_{i=1:n}$, n depending on the formulation that is chosen. In the Gelman-like formulation, we take six parameters (three parameters 15 controlling the preferences times 2 zooplankton types). In the spherical formulation, we take four parameters (two parameters controlling the preferences times two zooplankton types).

20 The ensemble contains 100 members. The background state ensemble is generated by adding a truncated-Gaussian perturbation to the solution $\mathbf{x}(t = 0)$.

$$\forall i = 1 : 100, \quad \mathbf{x}_b^i = \max(0, \mathbf{x}(t = 0) \times (1 + b_i)) \quad (11)$$

with $b \sim \mathcal{N}(0, \sigma_b^2)$. σ_b is chosen to be equal to 0.3 for all the state variables. In the Gelman-like formulation, the parameter ensemble is initialized by assuming that the parameters $(\phi_i)_{i=1:3}$ are normally distributed according to $\mathcal{N}(0, \sigma = 2)$. In the spherical 25 formulation, we assume that the parameters $(\phi_i)_{i=1:2}$ follow a triangular distribution:

$$\forall i = 1 : 2, \phi_i \sim \mathcal{T}(0, 1, \theta_i). \quad (12)$$

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



with $\theta_i \in [0, 1]$ the mode of the distribution. The probability density function reads:

$$\forall i = 1 : 2, f_{\phi_i}(\phi) = \begin{cases} \frac{2\phi}{\theta_i}, & \text{for } 0 \leq \phi \leq \theta_i \\ \frac{2(1-\phi)}{1-\theta_i}, & \text{for } \theta_i \leq \phi \leq 1 \end{cases} \quad (13)$$

The triangular distribution is simulated from the uniform distribution thanks to the MINMAX method suggested by Stein and KEBLIS (2009).

The prior values for the parameters $(\pi_i)_{i=1:3}$ are obtained by an adequate tuning of the 2 modes $(\theta_i)_{i=1:2}$. Equal preferences are obtained by solving the two nonlinear systems:

$$(\mathcal{S}_1) \quad \frac{\cos(\pi\theta_1) + 2\theta_1 - 1}{\pi^2\theta_1(1-\theta_1)} + \frac{1}{6} = 0. \quad (14)$$

$$(\mathcal{S}_2) \quad \frac{\cos(\pi\theta_2) + 2\theta_2 - 1}{\pi^2\theta_2(1-\theta_2)} = 0.$$

Solutions to these systems $(\mathcal{S}_i)_{i=1:2}$ exist (see Appendix B): θ_1 is obtained thanks to the MATLAB function *fzero* and is equal to 0.8905 and θ_2 is equal to 0.5. The mapping

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



of the preferences in Eq. (5) is as follows:

$$\begin{array}{cc}
 \text{Mesozooplankton} & \text{Microzooplankton} \\
 \left\{ \begin{array}{l} \pi_{\text{DIA}} = \cos^2\left(\frac{\pi}{2}\phi_1\right) \\ \pi_{\text{MIC}} = \sin^2\left(\frac{\pi}{2}\phi_1\right) \cos^2\left(\frac{\pi}{2}\phi_2\right) \\ \pi_{\text{DET}} = \sin^2\left(\frac{\pi}{2}\phi_1\right) \sin^2\left(\frac{\pi}{2}\phi_2\right) \end{array} \right. & \left\{ \begin{array}{l} \pi_{\text{DET}} = \cos^2\left(\frac{\pi}{2}\phi_1\right) \\ \pi_{\text{FLA}} = \sin^2\left(\frac{\pi}{2}\phi_1\right) \cos^2\left(\frac{\pi}{2}\phi_2\right) \\ \pi_{\text{DIA}} = \sin^2\left(\frac{\pi}{2}\phi_1\right) \sin^2\left(\frac{\pi}{2}\phi_2\right) \end{array} \right. \quad (15)
 \end{array}$$

This choice is motivated by our wish to respect the symmetrical relation between the two classes of phytoplankton in the definition of the observed variable (same weight for the FLA and DIA variables when computing the chlorophyll concentration). Since the mesozooplankton feeds the diatoms only, π_1 is equal to π_{DIA} . The microzooplankton feeding the two classes of phytoplankton, we choose π_2 and π_3 for representing π_{DIA} and π_{FLA} . Because of the asymmetric formulation of the parameterization, this choice of modeling can impact the estimation process: the estimates of the preferences obtained by the assimilation with the spherical formulation may depend on these choices. This is not true for the Gelman-like formulation.

For both formulations, anamorphosis functions are introduced to transform the state variables. These functions are also introduced to transform the parameters $(\phi_i)_{i=1:2}$ for the spherical formulation. However, the parameters $(\phi_i)_{i=1:3}$ already being normal for the Gelman-like formulation, their transformation by the anamorphosis functions is not necessary (see Sect. 2). The strategy to build the anamorphosis functions differs between the chlorophyll and the other state variables and parameters (if necessary) and is a variation of the hybrid approach described in Simon and Bertino (2012). Since the chlorophyll concentration in the ocean is usually assumed to have a log-normal distribution (Campbell, 1995), its anamorphosis function is the logarithmic function. For the other state variables and the parameters, the anamorphosis functions are built from

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

the empirical marginal distributions of the variables. The empirical anamorphosis functions are computed from a sample of the forecast ensemble and then are piecewise linearly interpolated to obtain the Gaussian anamorphosis functions. Their tails are linear and their last segments extrapolated towards specified biological minimum and maximum values. The spherical formulation introduces parameters that are bounded on both sides and for which the odds to reach the bounds during the assimilation are not null. The succession of analysis steps can build up discontinuities (“atoms”) of the distribution at the bounds which are not handled by the piecewise linear anamorphosis function – zero slopes are not invertible – (Simon and Bertino, 2012). Extending the first and last segments until they include the first values outside of the atoms seems to resolve the issue. The observation error ϵ^o is assumed to have a log-normal distribution: $\log(\epsilon^o) \sim \mathcal{N}(0, \sigma_o^2)$ with $\sigma_o = 0.3$. It results in a normal distributed observation error for the transformed observations with a standard deviation equal to 0.3.

Starting from this background, a one-year ensemble simulation is performed. The model includes truncated-Gaussian random perturbations on the phyto- and zooplankton components of the state variables every twelve hours. The standard deviation decreases linearly in the eight deepest layers in order to obtain a smooth transition between the deep layers and the bottom layer. No perturbations are added to the parameters and they remain constant during the model integration. Assimilation cycles are then performed over four years with a frequency of one analysis step every seven days. This frequency for observing the system is relatively low considering the short time scales of the bloom phenomenon. Figure 2 represents the evolution with time of the chlorophyll concentration in the two first surface layers in the reference solution and in the assimilated observations. We note during the blooms that the 7-day sampling of the reference leads to only one observation during the first peak in the concentration corresponding to the bloom of diatoms and only one or two observations during the second peak corresponding to the bloom of flagellates. Furthermore, the maximum values reached by the concentrations in the reference solution during these two peaks are generally not captured by the observations. So the blooms are mostly represented in

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

the observations by two Dirac pulses for which the amplitude and the timing are highly uncertain. This will result in difficulties for the ensemble-based Kalman filter methods to correctly estimate the state of the system, and notably some parameters. This is a real issue in the framework of ocean ecosystems due to the dynamics of the models – the weak production (apart from the bloom periods) results in low innovations and spread of the chlorophyll concentration in the ensemble, and so weak corrections by the filters during most of the year. An increase of the sampling frequency to four days would be enough to obtain a good representation of the blooms in this simple 1-D configuration, more specifically the transition phases, and potentially improve the quality of the estimation. Nevertheless, assimilating observations more frequently might not be affordable in real large configurations due to the computational costs that it implies. The use of an asynchronous version of the EnKF (Sakov et al., 2010) would be a solution to tackle these issues but is out of the scope of this study.

In order to check the robustness of the estimation against random initial conditions and observation errors, we repeated the experiment twenty times. That is, twenty initial ensembles (combined state-parameter background) and twenty sets of observations were generated. Nevertheless, the different assimilation systems used the same state component of the background ensemble and observations for each of the twenty realizations. The diagnostics shown in Sect. 4 are averaged over these twenty experiments.

4 Data assimilation results

4.1 Overall error evolution

We are interested in the time evolution of the relative Root Mean Square error (RMS) and the relative ensemble standard deviations (STD) of the solution of the two different formulations. These diagnostics are averaged over 20 experiments. The expression at

time t_n of these two quantities is as follows:

$$\text{RMS}(t_n) = \frac{\frac{1}{N_{\text{exp}}} \sum_{i=1}^{N_{\text{exp}}} \sqrt{\sum_{\mathbf{k} \in \Omega} (\mathbf{x}^t(t_n, \mathbf{k}) - \bar{\mathbf{x}}(t_n, \mathbf{k}, i))^2}}{\sqrt{\sum_{\mathbf{k} \in \Omega} \mathbf{x}^t(t_n, \mathbf{k})^2}} \quad (16)$$

$$\text{STD}(t_n) = \frac{\frac{1}{N_{\text{exp}}} \sum_{i=1}^{N_{\text{exp}}} \sqrt{\frac{1}{N-1} \sum_{\mathbf{k} \in \Omega} \sum_{m=1}^N (\mathbf{x}^m(t_n, \mathbf{k}, i) - \bar{\mathbf{x}}(t_n, \mathbf{k}, i))^2}}{\sqrt{\sum_{\mathbf{k} \in \Omega} \mathbf{x}^t(t_n, \mathbf{k})^2}} \quad (17)$$

where Ω is the domain of computation, N is the number of members, \mathbf{x}^m is the forecast member m , N_{exp} is the number of experiments, \mathbf{x}^t is the true state, and $\bar{\mathbf{x}}$ is the mean of the forecast ensemble.

Figure 3 represents the evolution of the relative RMS and standard deviation over five years for the diatoms, flagellates and the micro- and mesozooplankton. These diagnostics are average over the whole water column and Ω represents the 55 vertical layers. The evolution of the spatial average of the true state is plotted (green dashed line) in order to provide information on the yearly dynamics of the variables. No assimilation is performed during the first year. First, we note that both formulations lead to a reduction of the RMS error and the standard deviation for flagellates and their grazers, the microzooplankton. The peaks in the error for flagellates occur at the end of the flagellates blooms, which are too short in the assimilated solution notably around -25 m. The evolutions of the standard deviation and RMS error are in agreement during the last bloom both for microzooplankton and flagellates, which highlights a good

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



representation of the error by the ensemble during that period. An improvement of the detritus component of the solutions is also observed (not shown).

The impact of the data assimilation on the diatoms is mixed. We note a large increase of the standard deviation after all the diatoms blooms and a large peak in the RMS error during the first year with assimilation associated with a too long bloom. Then, the RMS error decreases year after year for both formulations and reaches its lowest values during the fourth year. However, a large peak is still present in the error during the final bloom for the spherical formulation. This is due to large values in the concentrations of diatoms localized around -70 m. Because the silica cycle depends only on the diatom concentration, these large peaks in the error result in a low increase of the RMS error for both silicate components during the bloom every year leading to final error around 10 % (not shown). In the same way, the assimilation cannot significantly reduce the RMS errors for the mesozooplankton. On average, the solutions obtained with the Gelman formulation present a lower error than the ones obtained with the spherical formulation. Finally, the nitrate and phosphate are not significantly impacted during the assimilation (not shown). On average, their RMS errors are low (less than 5 %) and exhibit low oscillations during the blooms.

4.2 Evolution of the parameters

Figure 4 represents the evolution with time of the mean and standard deviation of the ensemble for the meso- and microzooplankton grazing preferences. First, we note that both formulations lead on average to reasonably good final estimates of the microzooplankton grazing preferences. The largest corrections occurring during the first two blooms result in a convergence of the estimation towards the true values of the preferences in less than two years for both formulations. However, we note larger corrections during the last bloom with the Gelman formulation that can be explained by a larger spread for the preferences in the ensemble inherited from the initial ensembles. The Gelman formulation introduces a distribution with two parameters – the mean and the variance of the normal distribution, see Eq. (4) – which makes it possible to choose

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the mean and the standard deviation for the prior preferences. Our use of a distribution with one parameter – the mode of the triangular distribution, see Eq. (12) – allows only for the choice of the mean for the prior preferences. In these experiments, the prior variances chosen for the normal distributed parameters in the Gelman formulation leads to initial variances for the prior preferences larger than the ones obtained with the spherical formulation.

The mean and standard deviation of the twenty means of the preferences in the ensemble obtained at the end of the experiments are specified in Table 1. On average, the Gelman formulation produces slightly better estimates of the preferences for diatoms and detritus while both formulations lead to the same estimate of the preferences for flagellates. However, the ternary plots of the final estimates of the preferences for the twenty experiments in Fig. 5 show that the number of experiments for which the assimilation provides corrections in the direction of the true value for the three preferences is larger with the spherical formulation than with Gelman's: only two points do not belong to the shaded area representing the subspace of preferences defines by $0 \leq \pi_{\text{DET}} \leq \frac{1}{3}$, $\frac{1}{3} \leq \pi_{\text{FLA}} \leq 1$ and $0 \leq \pi_{\text{DIA}} \leq \frac{1}{3}$ (decrease of the preferences for the diatoms and detritus and increase of the preference for the flagellates) with the spherical formulation compared to four points with the Gelman formulation.

The estimation of the mesozooplankton grazing preferences is less successful. On average, we note in Fig. 4 that the corrections are very weak during the first two years of assimilation. The reduction of the standard deviation of the three preferences is very low for both formulations suggesting a weaker sensitivity of the surface chlorophyll to the mesozooplankton grazing preferences compared to the microzooplankton grazing preferences. This is highlighted in Fig. 6 by the Pearson correlation coefficients between the surface chlorophyll and the microzooplankton that are much larger than the ones between the surface chlorophyll and the mesozooplankton.

On average, the spherical formulation leads to significantly better final estimates of the preferences than the Gelman formulation (see Table 1). For this formulation, the

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

assimilation tends to correct strongly the preference for the detritus to the detriment of the preference for the microzooplankton. The ternary plots in Fig. 5 show that the estimation with the Gelman's formulation does not jointly improve the three preferences in 45 % of the experiments. For most of these experiments, this is due to an erroneous increase of the preference for the microzooplankton. The rate of failure decreases to 30 % of the experiments with the spherical formulation. For most cases, this is due to an erroneous increase of the preference for the detritus to the detriment of the preference for the diatoms. However, we think that these difficulties faced by the DENKF to correctly estimate the mesozooplankton grazing preferences are related to the framework of the experiments rather than the formulations of the change of variables. As stated earlier, the surface chlorophyll seems to be more sensitive to the microzooplankton than to the mesozooplankton in the model. Furthermore, improvements could be obtained by changing the experimental framework, for example the observation frequency, the specified observation error, etc.

5 Conclusions

In this study, we investigated the problem of estimating N positive sum-to-one constraint parameters with ensemble-based Kalman filters in the purpose of estimating zooplankton grazing preferences that are commonly used in ocean ecosystem models.

We have suggested a new formulation of the grazing preferences introducing a change of variables based on the hyperspherical coordinate system. This formulation results in the estimation of a reduced number ($N - 1$) of independent bounded parameters. Issues raised by estimating non-Gaussian distributed parameters with Kalman filters can be tackled by using the Gaussian anamorphosis. Furthermore, the nonlinear system of $N - 1$ equations to be solved in order to obtain equal prior preferences are also exhibited.

The performances of this approach and the one suggested by Gelman (1995) based on the Dirichlet distribution have been assessed in the framework of twin experiments

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



realized in a 1-D configuration of the coupled model GOTM-NORWECOM. Both approaches lead to correct estimates of the microzooplankton grazing preferences. They present the same difficulties to estimate the mesozooplankton grazing preferences that can be explained by the configuration of the experiments: the observed variable, the chlorophyll, constitutes only one type of food (diatoms) for the mesozooplankton diet compared to two (diatoms and flagellates) for the microzooplankton diet. Furthermore, the better results obtained with the spherical formulation for the mesozooplankton are not significant and cannot be guaranteed for more complex realistic configurations.

Both approaches present theoretical and practical advantages. The Gelman formulation leads to the estimation of Gaussian distributed parameters, a property that presents theoretical advantages in the context of Kalman filtering. This formulation is straightforward to apply for any number of preferences. However, it can require to estimate a large number of parameters in complex systems. The spherical formulation reduces the number of parameters to estimate but requires a choice of their prior distribution and to solve a nonlinear system of equations accordingly. In this study, we have used the triangular distribution for its simplicity and its applicability in our ecosystem model. But this distribution is not suitable for more than three preferences. From five preferences – N equal four can be solved via the introduction of the Hopf coordinate system – the questions of the choice of the distribution and the resolution of the system remains open. This suggests that the Gelman's formulation is more suitable in the framework of few zooplankton species with a diet involving numerous types of food while the spherical formulation could be more suitable in the framework of numerous zooplankton species with a diet involving few types of food.

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Appendix A

Derivation of the systems $(S_i)_{i=1:N-1}$ to obtain equal preferences $(\pi_i)_{i=1:N}$

Let $(\phi_i)_{i=1:N-1}$ be $N - 1$ independent random variables following the same distribution \mathcal{D} involving a set of parameters $(\Theta_i)_{i=1:N-1}$ that may have different values depending on i and with a support equal to the segment line $[0, 1]$. We note f_{ϕ_i} the probability density function of the parameter ϕ_i for all $i = 1:N - 1$:

$$\forall i = 1:N - 1, \quad f_{\phi_i} : [0, 1] \rightarrow \mathbb{R}_+ \\ \phi \mapsto f_{\phi_i}(\phi) \quad (\text{A1})$$

Let $(\pi_i)_{i=1:N}$ N random variables defined by the Eq. (5). We aim to choose the values of the set of parameters $(\Theta_i)_{i=1:N-1}$ to obtain equal expected values for the variables $(\pi_i)_{i=1:N}$:

$$\forall i = 1:N, \quad E[\pi_i] = \frac{1}{N} \quad (\text{A2})$$

First, let's start with a preliminary calculus. By integrating by parts and using the property of a probability density function $f: \int_{\mathbb{R}} f(\phi) d\phi = 1$, one has:

$$\int_0^1 \cos^2\left(\frac{\pi}{2}\phi\right) f_{\phi_i}(\phi) d\phi = \frac{1}{2} + \frac{1}{2} \int_0^1 \cos(\pi\phi) f_{\phi_i}(\phi) d\phi \quad (\text{A3})$$

By introducing the relation $\cos(a) = \frac{e^{ja} + e^{-ja}}{2}$ with $j^2 = -1$, it leads to:

$$\begin{aligned} \int_0^1 \cos^2\left(\frac{\pi}{2}\phi\right) f_{\phi_i}(\phi) d\phi &= \frac{1}{2} + \frac{1}{4} \int_0^1 e^{j\pi\phi} f_{\phi_i}(\phi) d\phi \\ &+ \frac{1}{4} \int_0^1 e^{-j\pi\phi} f_{\phi_i}(\phi) d\phi \\ &= \frac{1}{2} + \frac{1}{4} (\Phi_{\phi_i}(\pi) + \Phi_{\phi_i}(-\pi)) \end{aligned} \quad (\text{A4})$$

with Φ_{ϕ_i} the characteristic function of the parameter ϕ_i . By defining

$$\forall i = 1:N-1, \quad h(\Theta_i) = \frac{1}{4} (\Phi_{\phi_i}(\pi) + \Phi_{\phi_i}(-\pi)) \quad (\text{A5})$$

5 the last equation reads:

$$\int_0^1 \cos^2\left(\frac{\pi}{2}\phi\right) f_{\phi_i}(\phi) d\phi = \frac{1}{2} + h(\Theta_i) \quad (\text{A6})$$

In the same way, one has:

$$\int_{\mathbb{R}} \sin^2\left(\frac{\pi}{2}\phi\right) f_{\phi_i}(\phi) d\phi = \frac{1}{2} - h(\Theta_i) \quad (\text{A7})$$

Now, for $i = 1$ one has:

$$10 \quad E[\pi_1] = \int_0^1 \cos^2\left(\frac{\pi}{2}\phi\right) f_{\phi_1}(\phi) d\phi = \frac{1}{2} + h(\Theta_1) \quad (\text{A8})$$

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

and the Eq. (A2) reads:

$$h(\Theta_1) = \frac{2 - N}{2N} \quad (\text{A9})$$

which is equivalent to the system (S_1) defined by the Eq. (6).

Now, let i be an integer between 2 and $N - 1$. By definition of the variables $(\pi_i)_{i=1:N}$ one has:

$$\forall i = 2:N - 1, \quad E[\pi_i] = \int_{\mathbb{R}^{N-1}} \prod_{k=1}^{i-1} \sin^2\left(\frac{\pi}{2}\phi_k\right) \cos^2\left(\frac{\pi}{2}\phi_i\right) f_{(\phi_i)_{i=1:N-1}}(\phi) d\phi \quad (\text{A10})$$

Because the variables $(\phi_i)_{i=1:N-1}$ are independent, it leads to:

$$\forall i = 2:N - 1, \quad E[\pi_i] = \prod_{k=1}^{i-1} \int_{\mathbb{R}} \sin^2\left(\frac{\pi}{2}\phi_k\right) f_{\phi_k}(\phi) d\phi \int_{\mathbb{R}} \cos^2\left(\frac{\pi}{2}\phi_i\right) f_{\phi_i}(\phi) d\phi \quad (\text{A11})$$

By introducing the Eqs. (A6) and (A7), one obtains:

$$\forall i = 2:N - 1, \quad E[\pi_i] = \prod_{k=1}^{i-1} \left(\frac{1}{2} - h(\Theta_k)\right) \left(h(\Theta_i) + \frac{1}{2}\right) \quad (\text{A12})$$

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



It leads to, $\forall i = 2 : N - 1$:

$$\begin{aligned}
 \frac{E[\pi_i]}{E[\pi_{i-1}]} = 1 &\Leftrightarrow \frac{\prod_{k=1}^{i-1} \left(\frac{1}{2} - h(\Theta_k)\right) \left(h(\Theta_i) + \frac{1}{2}\right)}{\prod_{k=1}^{i-2} \left(\frac{1}{2} - h(\Theta_k)\right) \left(h(\Theta_{i-1}) + \frac{1}{2}\right)} = 1 \\
 &\Leftrightarrow \frac{\left(\frac{1}{2} - h(\Theta_{i-1})\right) \left(h(\Theta_i) + \frac{1}{2}\right)}{\left(h(\Theta_{i-1}) + \frac{1}{2}\right)} = 1 \\
 &\Leftrightarrow h(\Theta_i) = -\frac{1}{2} + \frac{1 + 2h(\Theta_{i-1})}{1 - 2h(\Theta_{i-1})}
 \end{aligned} \tag{A13}$$

Finally, we obtain a recurrence relation between the variables $(h(\Theta_i))_{i=1:N-1}$:

$$\left\{ \begin{array}{l} h(\Theta_1) = \frac{2 - N}{2N} \\ \forall i = 2:N - 1, \\ h(\Theta_i) = \frac{1}{2} + \frac{1 + 2h(\Theta_{i-1})}{1 - 2h(\Theta_{i-1})} \end{array} \right. \tag{A14}$$

5 The solution of Eq. (A14) is given by:

$$\forall i = 1:N - 1, \quad h(\Theta_i) = -\frac{N - i - 1}{2(N - i + 1)} \tag{A15}$$

which corresponds to the systems $(S_i)_{i=1:N-1}$.

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

We must now check that the relation $E[\pi_{N-1}] = E[\pi_N]$ is satisfied.

$$\begin{aligned} \frac{E[\pi_N]}{E[\pi_{N-1}]} = 1 &\Leftrightarrow \frac{\int_{\mathbb{R}} \sin^2\left(\frac{\pi}{2}\phi_{N-1}\right) f_{\phi_{N-1}}(\phi) d\phi}{\int_{\mathbb{R}} \cos^2\left(\frac{\pi}{2}\phi_{N-1}\right) f_{\phi_{N-1}}(\phi) d\phi} = 1 \\ &\Leftrightarrow \frac{1 - 2h(\Theta_{N-1})}{1 + 2h(\Theta_{N-1})} = 1 \\ &\Leftrightarrow h(\Theta_{N-1}) = 0 \end{aligned} \quad (\text{A16})$$

Applying Eq. (A15) for $i = N - 1$ leads to $h(\Theta_{N-1}) = 0$. So, one does have $E[\pi_{N-1}] = E[\pi_N]$.

5 Appendix B

Triangular distribution: on the existence of solutions for the systems $(S_i)_{i=1:N-1}$

We aim to exhibit the conditions for the existence of a solution to the $N - 1$ systems $(S_i)_{i=1:N-1}$:

$$\forall i = 1:N - 1,$$

$$(S_i) \quad \frac{1}{4}(\Phi_{\phi_i}(\pi) + \Phi_{\phi_i}(-\pi)) + \frac{N - i - 1}{2(N - i + 1)} = 0. \quad (\text{B1})$$

10 with $(\phi_i)_{i=1:N-1}$ triangular distributed: $\forall i = 1 : N - 1, \phi_i \sim \mathcal{T}(0, 1, \theta_i)$. For each parameter, the characteristic function Φ_{ϕ_i} reads:

$$\forall i = 1:N - 1,$$

$$\forall t \in \mathbb{R}, \quad \Phi_{\phi_i}(t) = -2 \frac{(1 - \theta_i) - e^{j\theta_i t} + \theta_i e^{jt}}{\pi^2 \theta_i (1 - \theta_i)}. \quad (\text{B2})$$

So, the systems $(S_i)_{i=1:N-1}$ read:

$$\forall i = 1:N - 1,$$

$$(S_i) \quad \frac{\cos(\pi\theta_i) + 2\theta_i - 1}{\pi^2\theta_i(1 - \theta_i)} + \frac{N - i - 1}{2(N - i + 1)} = 0. \quad (B3)$$

For $i = N - 1$, $\theta_{N-1} = \frac{1}{2}$ is a trivial solution to

$$\cos(\pi\theta_{N-1}) + 2\theta_{N-1} - 1 = 0. \quad (B4)$$

5 Let i be an integer between 1 and $N - 2$. Furthermore, we note $a_N^i = \frac{N - i - 1}{2(N - i + 1)}$.

So, (S_i) read

$$(S_i) \quad \frac{\cos(\pi\theta_i) + 2\theta_i - 1}{\pi^2\theta_i(1 - \theta_i)} + a_N^i = 0 \quad (B5)$$

$$\Leftrightarrow \cos(\pi\theta_i) = \pi^2 a_N^i \theta_i^2 - (2 + \pi^2 a_N^i) \theta_i + 1$$

$$\Leftrightarrow \cos(\pi\theta_i) = P_N^i(\theta_i)$$

with $P_N^i(\theta) = \pi^2 a_N^i \theta^2 - (2 + \pi^2 a_N^i) \theta + 1$ a polynomial of degree 2 (a_N^i is non null since $i \leq N - 2$).

10 The discriminant Δ_N^i of P_N^i is equal to $\Delta_N^i = 4 + \pi^4 (a_N^i)^2$ and is definite positive. P_N^i

has two distinct roots $\theta_i^\pm = \frac{2 + a_N^i \pi^2 \pm \sqrt{4 + (a_N^i)^2 \pi^4}}{2a_N^i \pi^2}$. The polynomial function $\theta \rightarrow$

$P_N^i(\theta)$ has a minimum ($a_N^i > 0$) for $\tilde{\theta}_i = \frac{1}{2} + \frac{1}{a_N^i \pi^2}$. Furthermore, one has $P_N^i(0) = 1 =$

$\cos(0)$ and $P_N^i(1) = -1 = \cos(\pi)$.

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



It can be shown that the existence of a solution θ_i , $0 < \theta_i < 1$, for the system (S_i) will depend on the value of $\tilde{\theta}_i$ and that $\theta_i > \frac{1}{2}$ if existence. For the sake of simplicity, we do not exhibit the full proof, but rather highlight the results. Two cases have to be discriminated. On the one hand, if $\tilde{\theta}_i < 1$, i.e. $a_N^i > a^* = \frac{2}{\pi^2}$, $\forall \theta \in]0, 1[$, $P_N^i(\theta) < \cos(\pi\theta)$ and the system (S_i) does not admit solutions. This case is illustrated on Fig. B1 by the blue and green curves that do not intersect on the segment line $]0, 1[$, with 0 and 1 excluded. On the other hand, if $\tilde{\theta}_i > 1$, the system (S_i) admits a solution between $\frac{1}{2}$ and 1. This case is illustrated on Fig. B1 by the red and green curves.

A solution of the system (S_i) will be numerically found if and only if $a_N^i > \frac{2}{\pi^2}$ that is equivalent to:

$$N - i < \frac{\pi^2 + 4}{\pi^2 - 4} \sim 2.36 \quad (\text{B6})$$

It means that only the systems (S_{N-1}) and (S_{N-2}) admit a solution. In practice, it means that for more than 3 preferences ($N \geq 4$), it will not be possible to obtain a prior with equal preferences when using triangular distributed parameters $(\phi_i)_{i=1:N-1}$.

Acknowledgements. This study has been funded by the eVITA-EnKF project from the Research Council of Norway and the MyOcean and GreenSeas IP from the European Commission's 7th FP. A grant of CPU time from the Norwegian Supercomputing Project (NOTUR2) has been used.

References

Aksnes, D., Ulvestad, K., Baliño, B., Berntsen, J., and Svendsen, E.: Ecological modelling in coastal waters: towards predictive physical-chemical-biological simulation models, *Ophelia*, 41, 5–36, 1995. 1093

**Estimation of the
zooplankton grazing
preferences with the
DENKF**E. Simon et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

- Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, 129, 2884–2903, 2001. 1087
- Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, *Int. Stat. Rev.*, 71, 223–241, 2003. 1088
- 5 Bertino, L. and Lisæter, K. A.: The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans, *Journal of Operational Oceanography*, 1, 15–19, 2008. 1093
- Buitenhuis, E., Le Quéré, C., Aumont, O., Beaugrand, G., Bunker, A., Hirst, A., Ikeda, T., O'Brien, T., Piontkovski, S., and Straile, D.: Biogeochemical fluxes through mesozooplankton, *Global Biogeochem. Cy.*, 20, GB2003, doi:10.1029/2005GB002511, 2006. 1087
- 10 Buitenhuis, E. T., Rivkin, R. B., Saille, S., and Le Quéré, C.: Biogeochemical fluxes through microzooplankton, *Global Biogeochem. Cy.*, 24, GB4015, doi:10.1029/2009GB003601, 2010. 1087
- Burchard, H., Bolding, K., and Villareal, M. R.: GOTM, a general ocean turbulence model: theory, implementation and test cases, European Commission Technical Report, Brussels, 1999. 1093
- 15 Burchard, H., Deleersnijder, E., and Meister, A.: Application of modified Patankar schemes to stiff biogeochemical models for the water column, *Ocean Dynam.*, 55, 326–337, 2005. 1093
- Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, 100, 13237–13254, 1995. 1097
- 20 Doron, M., Brasseur, P., and Brankart, J.-M.: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model: twin experiments, *J. Marine Syst.*, 87, 3–4, 2011. 1088
- Evans, G. T. and Parslow, J. S.: A model of annual plankton cycles, *Biological Oceanography*, 3, 327–347, 1985. 1086
- 25 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10182, 1994. 1087
- Evensen, G.: The Ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003. 1087, 1089
- 30 Evensen, G.: The Ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, 29, 83–104, 2009. 1087
- Gelman, A.: Method of Moments Using Monte Carlo Simulation, *Journal of Computational and Graphical Statistics*, 4, 36–54, 1995. 1088, 1090, 1103

**Estimation of the
zooplankton grazing
preferences with the
DEnKF**

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- Gelman, A., Bois, F., and Jiang, J.: Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions, *J. Am. Stat. Assoc.*, 91, 1400–1412, 1996. 1088
- Gentleman, W., Leising, A., Frost, B., Strom, S., and Murray, J.: Functional responses for zooplankton feeding on multiple resources: a review of assumptions and biological dynamics, *Deep Sea Res. II*, 50, 2847–2875, 2003. 1086, 1087
- 5 Gregg, W. W. and Casey, N. W.: Global and regional evaluation of the SeaWiFS chlorophyll data set, *Remote Sens. Environ.*, 93, 463–479, 2004.
- Pätsch, J., Kühn, W., Moll, A., and Lenhart, H: ECOHAM4 user guide, Technical Report 1-2009, Institut für Meereskunde, Hamburg, Germany, 2009. 1093
- 10 Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filter, *Tellus*, 60A, 361–371, 2008.
- Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, *Tellus*, 62A, 24–29, 2010. 1099
- Simon, E. and Bertino, L.: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment, *Ocean Sci.*, 5, 495–510, doi:10.5194/os-5-495-2009, 2009. 1088
- 15 Simon, E. and Bertino, L.: Gaussian anamorphosis extension of the DEnKF for combined state and parameter estimation: application to a 1D ecosystem model, *J. Marine Syst.*, 89, 1–18, 2012. 1088, 1089, 1097, 1098
- 20 Skogen, M. and Søliland, H.: A user's guide to NORWECOM v2.0. The NORwegian Ecological Model system, Technical Report Fisken og Havet 18, Institute of Marine Research, Norway, 1998. 1093
- Stein, W. E. and Kiehl, M. F.: A new method to simulate the triangular distribution, *Mathematical and Computer Modelling*, 49, 1143–1147, 2009. 1096
- 25 Umlauf, L. and Burchard, H.: Second-order turbulence closure models for geophysical boundary layers. A review of recent work, *Cont. Shelf Res.*, 25, 795–827, 2005. 1093
- Zhou, H., Gómez-Hernández, J. J., Hendricks Franssen, H.-J., and Li, L.: An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Adv. Water Res.*, 34, 844–864, 2011. 1088

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Table 1. Zooplankton grazing preferences (π_i)_{*i*=1;3}: mean and standard deviation (computed over the twenty experiments) of the means of preferences obtained at the final time.

Mesozooplankton			
Diet	Diatoms	Microzooplankton	Detritus
True value	0.6	0.15	0.25
Gelman's formulation	0.50±0.19	0.31±0.16	0.19±0.09
Spherical formulation	0.51±0.19	0.24±0.11	0.25±0.13
Microzooplankton			
Diet	Detritus	Flagellates	Diatoms
True value	0.15	0.6	0.25
Gelman's formulation	0.19±0.1	0.56±0.09	0.25±0.05
Spherical formulation	0.20±0.09	0.56±0.10	0.24 ±0.05

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[⏪](#)
[⏩](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

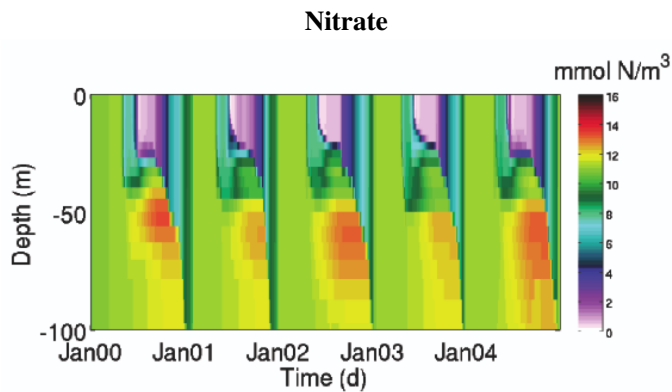
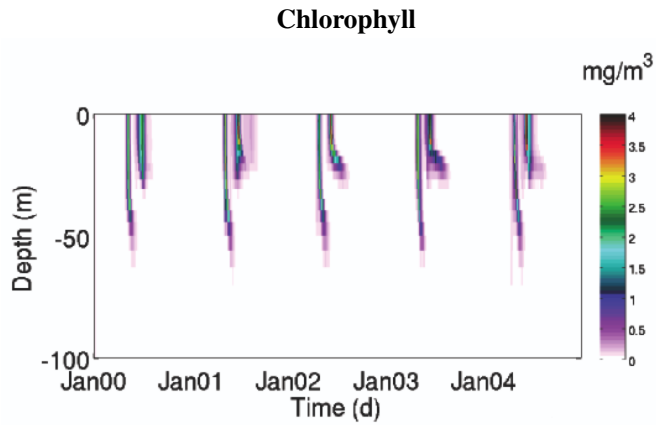


Fig. 1. Reference solution: time evolution of chlorophyll and nitrate in the upper 100 m from 1 January 2000 to 31 December 2004.

**Estimation of the
zooplankton grazing
preferences with the
DENKF**

E. Simon et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	⏩
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

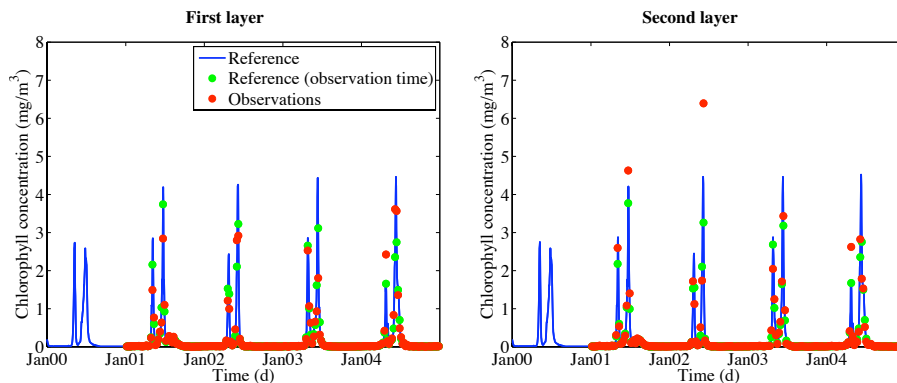


Fig. 2. Observations: evolution with time of the chlorophyll concentration at the surface (2 first layers) in the reference solution (blue line), in the reference solution at observation times (green circles) and in the observations (red circles) for one experiment.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

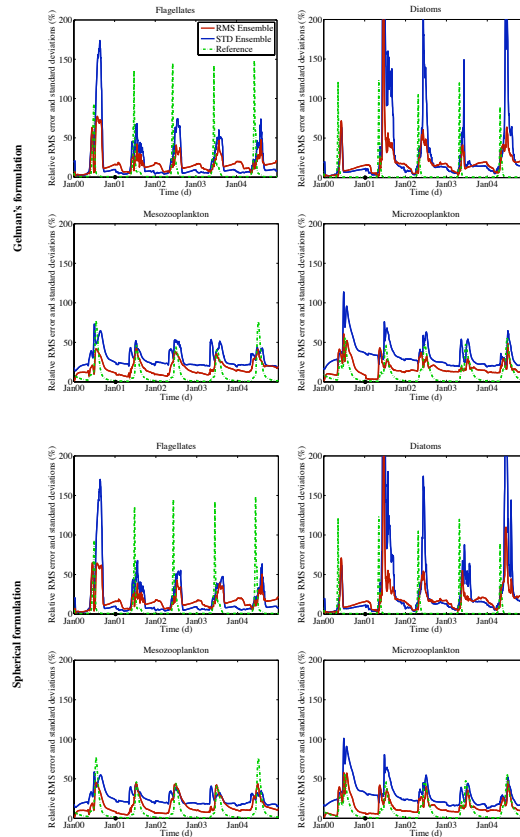


Fig. 3. Evolution with time of the relative RMS error and standard deviation computed over the water column and averaged over the twenty experiments. The spatial mean of the reference solution is plotted to highlight the seasonal dynamics (green dashed curve). The black dot highlights the date of the first analysis.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

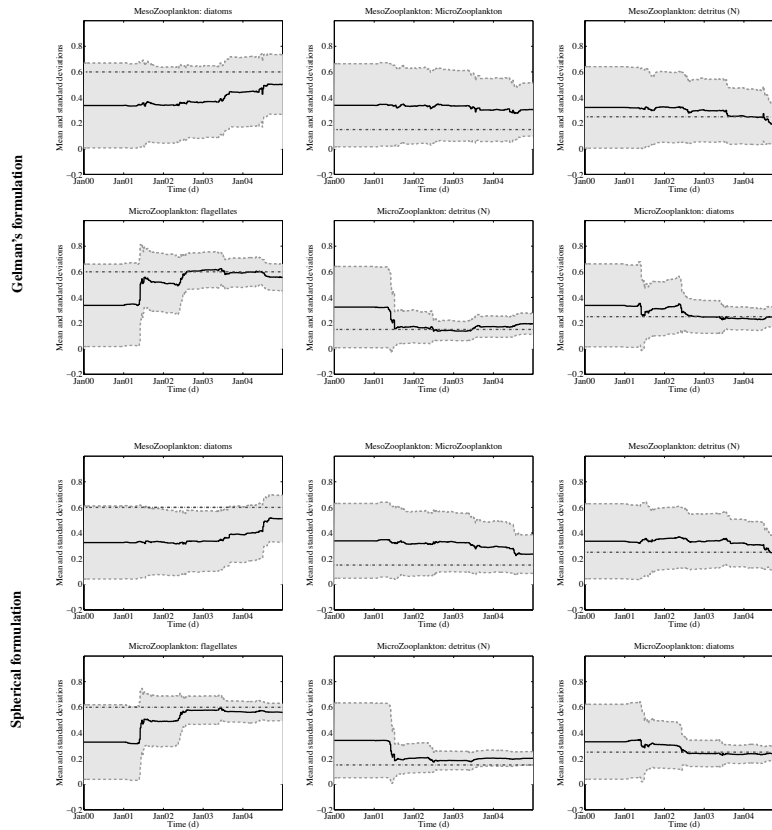


Fig. 4. Evolution with time of the averaged mean (black line) and averaged mean plus/minus the standard deviation (shaded area) of the grazing preferences. The true value is highlighted with a dark dash-dote line.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[⏪](#)
[⏩](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

Estimation of the zooplankton grazing preferences with the DEnKF

E. Simon et al.

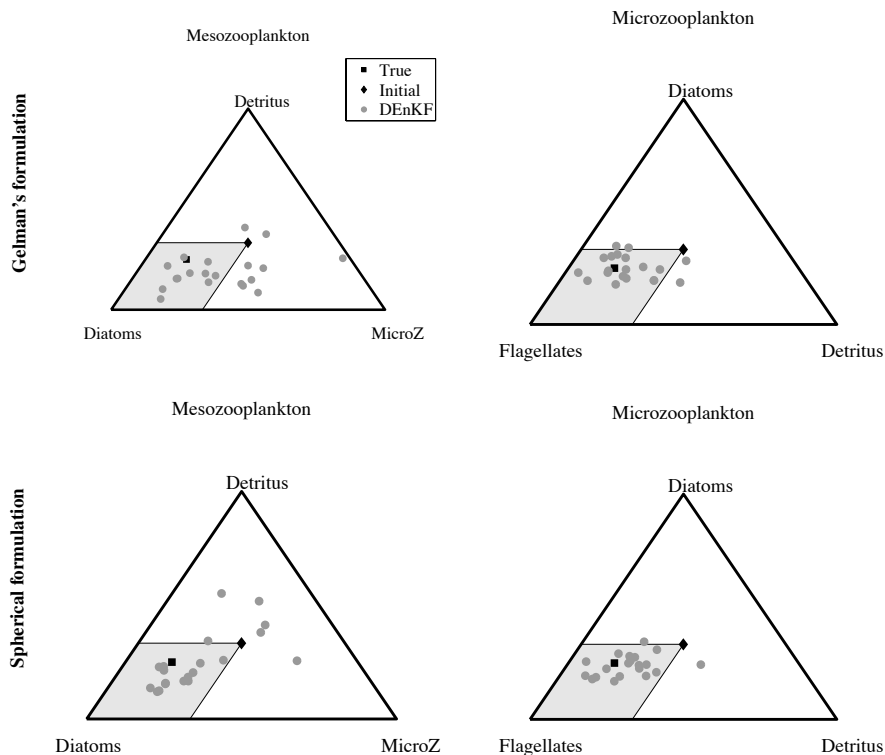


Fig. 5. Ternary plots of the final estimate (mean of the ensemble) of the grazing preferences parameters for the twenty experiments. The estimates obtained after assimilation are plotted with grey circles, the true set of parameters with a black square and the mean of the background set of parameters with a black diamond.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

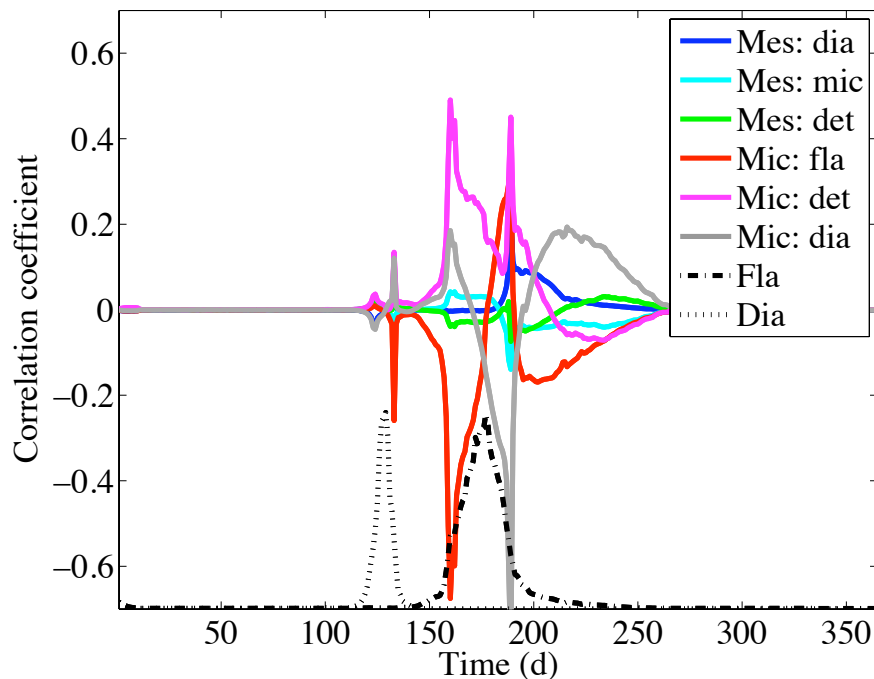


Fig. 6. Spherical formulation: evolution with time of the Pearson correlation coefficients between the surface chlorophyll and the meso- and microzooplankton grazing preferences during the first year (no assimilation) and averaged over the twenty experiments. The evolution of the averaged surface chlorophyll concentration due to diatoms (resp. flagellates) is plotted with a dark dotted line (resp. with a dark dash-dotted line).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

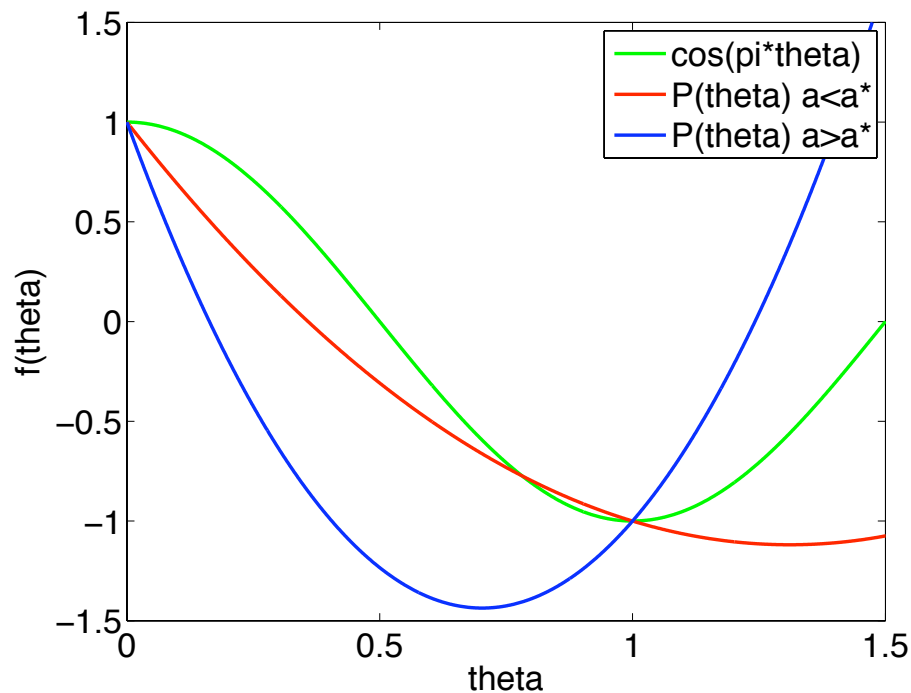


Fig. B1. Existence of a solution of the system $(S)_{i=1:N-1}$. In green: $\theta \rightarrow \cos(\pi\theta)$. In red: $\theta \rightarrow P_N^i(\theta)$ for $a_N^i < a^* = \frac{2}{\pi^2}$. In blue: $\theta \rightarrow P_N^i(\theta)$ for $a_N^i > a^* = \frac{2}{\pi^2}$.

Estimation of the zooplankton grazing preferences with the DENKF

E. Simon et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion