Journal of Marine Systems 89 (2012) 1–18

Contents lists available at ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys

Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model

Ehouarn Simon *, Laurent Bertino

Nansen Environmental and Remote Sensing Center, Thormøhlensgate 47, 5006 Bergen, Norway

ARTICLE INFO

Article history:
Received 13 December 2010
Received in revised form 24 June 2011
Accepted 11 July 2011
Available online 24 July 2011

Keywords:
Data assimilation
Ensemble Kalman filter
Gaussian anamorphosis
Combined state–parameter estimations
Ecosystem modeling

ABSTRACT

We consider the problem of combined state–parameter estimations in biased nonlinear models with non-Gaussian extensions of the Deterministic Ensemble Kalman Filter (DEnKF). We focus on the particular framework of ocean ecosystem models. Such models present important obstacles to the use of data assimilation methods based on Kalman filtering due to the non-linearity of the models, the constraints of positiveness that apply to the variables and parameters, and the non-Gaussian distribution of the variables in which they result.

We present extensions of the DEnKF dealing with these difficulties by introducing a nonlinear change of variables (anamorphosis function) in order to execute the analysis step with Gaussian transformed variables and parameters. Several strategies to build the anamorphosis functions are investigated and compared within the framework of twin experiments realized in a simple 1D ocean ecosystem model. A solution to the problem of the specification of the observation error for transformed observations is suggested. The study highlights the inability of the plain DEnKF with a simple post-processing of the negative values to properly estimate parameters when constraints of positiveness apply to the variables. It goes on to show that the introduction of the Gaussian anamorphosis can remedy these assimilation biases.

1. Introduction

The knowledge of the status of marine resources should be closely monitored in a changing climate, so analysis and short term forecasts of the primary production are needed by environmental agencies for monitoring algal blooms and possible movement of the fish populations (Allen et al., 2008; Johannessen et al., 2007). To that end, within the framework of the MyOcean[1] project, research activities attempt the reanalysis of the primary production and the biological components of the oceans, notably for the Arctic through the Arctic Marine Forecasting Center.

The numerical ecosystem models developed during the last decades, as well as their coupling with existing physical ocean models, are a necessary step, together with the direct observation of the ocean biology, to meet these goals. Nevertheless these models present numerous uncertainties linked to the complexities of the processes that they attempt to represent and the parameterizations that they introduce. Even though many improvements have been made in the modeling of ocean ecosystems, the models are still too simple in comparison to the complexity of the ocean biology. Parameters remain poorly known, cannot be observed (lack of direct measure) and may vary in space and in time (Losa et al., 2003; Losa et al., 2004). Furthermore, due to the key role played by the parameters or the mathematical model chosen for the parameterizations, wrong specifications of parameters can lead to large model error.

The data assimilation methods, thanks to their ability to combine in an optimal way the heterogeneous and uncertain information provided by the models and the observations, are relevant tools to tackle the problem of parameter calibration. The feasibility and the potentialities of simultaneous state and parameter estimations with ensemble-based Kalman filters have been demonstrated by Anderson (2001), who augmented the state vector with the parameters to estimate. In the same way, experiments of combined state–parameter estimations conducted in a simple linear scalar model (Evensen, 2006) highlighted the abilities of the Ensemble Kalman Filter (EnKF; Evensen (1994, 2003)) to calibrate a poorly known parameter. This approach has also proved to be efficient even for large scale applications, as highlighted by the twin experiments in an earth system model of intermediate complexity (Annan et al., 2005). Furthermore, the authors introduce logarithm transformations to guarantee the positiveness of several diffusion parameters during the estimation. The performance of the EnKF in combined state–parameter estimations have been carefully evaluated in a 2-D sea-breeze model by Aksoy et al. (2006). In the framework of ocean biogeochemistry, Losa et al. (2004) successfully applied weak constraint variational data assimilation to estimate parameters in a 0-D ecosystem model.

Nevertheless application of data assimilation methods to ecosystem models in an efficient way is a theoretically and practically challenging issue. On the one hand, the strongly nonlinear behavior of ecosystem models (especially during the period of the spring bloom) raises the question of which stochastic model to use (Bertino et al., 2003). On the

* Corresponding author. Tel.: +47 55205863; fax: +47 55205801.
E-mail addresses: ehouarn.simon@nersc.no (E. Simon), laurent.bertino@nersc.no (L. Bertino).

[1] http://www.myocean.eu.org.

0924-7963/$ – see front matter
doi:10.1016/j.jmarsys.2011.07.007

other hand, one is also confronted with the model constraints. Most variables of ecosystem models are concentrations of a biological tracer, and cannot be negative. In the same way, parameters are positive. These non-Gaussian distributions of most biogeochemical variables and parameters break an important assumption of the linear analysis, leading to a loss of optimality of the EnKF (and other linear filters). The optimality of the linear statistical analysis is proved under some assumptions, notably an assumption of Gaussianity made on the distribution of the variables (of the model and the observations) and the errors. Twin experiments of combined state–parameter estimations with the Lorenz model done by Kivman (2003) highlighted the difficulties — indeed the inability — of the EnKF to recover the true value of positive parameters in nonlinear models. This study confirmed also the benefits of using nonlinear methods like particle filters in such non-Gaussian frameworks. Unlike the EnKF, the application of an extension of a Sequential Importance Resampling filter (SIR; see Doucet et al., 2001) led to successful estimations of the true value of parameters. In the same way, Losa et al. (2003) successfully applied a SIR filter for combined state–parameter estimations in a 1-D ecosystem model. Indeed, particle filters seem attractive for such models as they are variance minimizing schemes for any probability density function. However, the size of the ensemble required for an efficient application of such a filter is too large to be considered for realistic configurations (Snyder et al., 2008). We refer to the review of van Leeuwen (2009) for more details about particle filters.

An easy method of performing Kalman filter estimation in an extended framework of variables with non-Gaussian distributions involves the introduction of the Gaussian anamorphosis into the filter, as suggested by Bertino et al. (2003). The idea is to introduce non-linear changes of variables (anamorphosis functions) in order to realize the analysis step with Gaussian distributed transformed variables. The numerical experiments of model state estimation that they conducted in a 1-D ocean ecosystem model highlighted the potential of this approach. In a previous study (Simon and Bertino, 2009), we demonstrated that this non-Gaussian extension of the EnKF could be easily applied to perform model state estimations in realistic configurations. Twin experiments done in a 3-D configuration of the North Atlantic and Arctic oceans highlighted a slight advantage in effectiveness compared to the plain EnKF with post-processing of the negative values. This advantage has been recently confirmed by Doron et al. (2011) in the framework of twin experiments of combined state–parameter estimations in a 3-D ocean-coupled physical–biogeochemical model. The unique analysis performed during the spring bloom highlighted the ability of this non-Gaussian extension of the Kalman filter “to retrieve consistently the maps of parameters and thus reduce their prior uncertainty”, as stated by the authors. Nevertheless, this study does not provide information about potential assimilation bias that might occur after several cycles of analysis.

The present study extends our previous research to the problem of combined state–parameter estimations in inevitably biased ecosystem models, and we focus on Kalman filtering. More information concerning the more general problem of Gaussian statistical modeling in data assimilation can be found in Bocquet et al. (2010). The aim of this study is to demonstrate that the Deterministic Ensemble Kalman filter (DEnKF; (Sakov and Oke, 2008)) — and more generally ensemble-based Kalman filters — remains a high-performance tool for the estimation of biased parameters in such non-Gaussian frameworks involving positive state variables and parameters, provided that the variables and parameters are appropriately transformed before and after the analysis. In that way, we focus on the strategies to empirically design the anamorphosis functions. Attention is also given to the problem of the specification of the observation error for the transformed variables.

The outline of this paper is as follows. We present a non-Gaussian extension of the DEnKF and different strategies to build the anamorphosis function in Section 2. We describe our experimental framework in Section 3, present and discuss our results in Section 4 and make our conclusions in Section 5.

2. Non-Gaussian extensions of the deterministic ensemble Kalman filter

We describe in this section a way to design a non-Gaussian extension of the DEnKF. Essentially we introduce nonlinear changes of variables in order to realize the analysis step with Gaussian distributed transformed variables, while the forecast step is done in the physical or biological space.

2.1. The deterministic ensemble Kalman filter with Gaussian anamorphosis

As suggested by Bertino et al. (2003) for the EnKF, the algorithm is based on the skeleton of the DEnKF and divides into two steps:

Forecast: the forecast step is a propagation step that uses a Monte-Carlo sampling to approximate the forecast density by N realizations:

$$\forall i = 1:N, \quad \mathbf{x}_n^{f,i} = f_{n-1}\left(\mathbf{x}_{n-1}^{a,i}, \epsilon_n^{m,i}\right) \tag{1}$$

with $\mathbf{x}_n$ the state vector at time t_n, f_{n-1} the nonlinear model and ϵ_n^m the model error. The superscripts ^{f} and ^{a} stand for forecast and analysis.

Analysis: the analysis step conditions each forecast member to the new observation $\mathbf{y}_n$ by a linear update. The anamorphosis functions are introduced in this step.

For each variable of the model, at time t_n, we apply a function ψ_n which is a nonlinear bijective function from the physical space to a Gaussian space. We transform each variable separately. In order to simplify the notations, we consider the monovariate case (i.e., there is only one function ψ_n). It reads:

$$\forall i = 1:N, \quad \tilde{\mathbf{x}}_n^{f,i} = \psi_n\left(\mathbf{x}_n^{f,i}\right). \tag{2}$$

In practice, this means that we apply a transformation to each variable in every point of the discretized domain.

In the same way, we introduce an anamorphosis function χ_n for the observations $\mathbf{y}_n$ at time t_n:

$$\tilde{\mathbf{y}}_n = \chi_n(\mathbf{y}_n). \tag{3}$$

The observation operator $\mathbf{H}$ links the physical variables and the observations. We define the observation operator $\tilde{\mathbf{H}}_n$ linking the transformed variables and observations by the formula

$$\tilde{\mathbf{H}}_n = \chi_n \circ \mathbf{H} \circ \psi_n^{-1} \tag{4}$$

where $\circ$ defines the function composition.

The linear analysis is done with the transformed variables and observations according to the equations for the updates of the mean and the ensemble anomalies of the DEnKF described in Sakov and Oke (2008). The transformed Kalman gain matrix is built on the forecast error covariance matrix $\tilde{\mathbf{C}}_n^f$ approximated by the covariance of $\left(\tilde{\mathbf{x}}_n^{f,i}\right)_{i=1:N}$.

The inverse transformation to the physical space is done by using the inverse of the anamorphosis function:

$$\forall i = 1:N, \quad \mathbf{x}_n^{a,i} = \psi_n^{-1}\left(\tilde{\mathbf{x}}_n^{a,i}\right). \tag{5}$$

The analyzed mean $\mathbf{x}_n^a$ and the covariance matrix $\mathbf{C}_n^a$ are approximated by the ensemble average and covariance of $\left(\mathbf{x}_n^{a,i}\right)_{i=1:N}$.

2.2. Strategies to design a monovariate anamorphosis function

The performances of the non-Gaussian extensions of the DEnKF described above are strongly dependent on the choice of the anamorphosis functions ψ_n and χ_n. One solution is to use analytic functions like the logarithm. However, this requires prior knowledge of the distribution of variables. Another solution is to construct the