

Biogeochemical Oceanographic Data Assimilation: Dimensionality Reduced Kalman Filter

For Mediterranean Sea Forecasting



S. Spada^{1,2}, G. Cossarini¹, S. Salon¹ & S. Maset²

¹Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Trieste, Italy

²Università degli studi di Trieste, Trieste, Italy

simonespada.info@gmail.com

sspada@inogs.it

Abstract

Data assimilation is a key element to improve the performance of biogeochemical ocean/marine forecasting systems. Handling the very big dimension of the state vector of the system (often of the order of 10^6) remains an issue, also considering the computational efficiency of operational systems. Indeed, simple product operations involving the covariance matrices are too heavy to be computed for operational forecasting purposes. Various attempts have been made in literature to reduce the complexity of this task, often adding strong hypotheses to simplify the problem and decrease the computational cost.

The MedBFM model system ([5], [8], and references thereby), which is responsible for monitoring and forecasting the biogeochemical state of the Mediterranean Sea within the European Copernicus Marine Services (see <http://marine.copernicus.eu/>) assimilates surface chlorophyll data through a 3D Variational algorithm, that decomposes the background error covariance matrix into sequential operators to reduce complexity [8].

In this work, we developed a novel Kalman Filter for the MedBFM system. The novel Kalman Filter scheme starts from a SEIK ([9]) approach but benefits from advanced Principal Component Analysis (PCA, [4]) to reduce the dimension of covariance matrices and improve the computational efficiency.

We compared the standard SEIK filter and the new Kalman filter implementations in a one dimensional transport model with 2 biological variables in terms of root mean square distance. In the vast majority of the experiments, the new Kalman filter had better performances.

The SEIK Filter

The notation mostly follows [7]

Background state:

Background error covariance:

Model error covariance:

Observation:

Observation error covariance:

Perturbation matrix:

Matrix of ensemble of N members:

Evolution of the ensemble:

Observation operator on the columns of X^f :

Error subspace basis:

Forecast

Forecast covariance:

Analysis covariance:

Analysis:

$$\begin{aligned}
 & x^b \in \mathbb{R}^n \\
 & P^b = LAL^T, L \in \mathbb{R}^{n \times N-1}, A \in \mathbb{R}^{N-1 \times N-1} \\
 & Q \\
 & y \\
 & R \\
 & Z = \sqrt{N}LC\Omega^T, C \in \mathbb{R}^{N-1 \times N-1}, \Omega \in \mathbb{R}^{N \times N-1}, \\
 & \text{with } CC^T = A \text{ and } \Omega \text{ random matrix} \\
 & \text{with orthonormal columns and zero columns sum} \\
 & X = x^b \mathbb{1}_{1 \times N} + Z \\
 & X^f \\
 & Y \\
 & L^f = X^f T, \\
 & \text{with } T = \begin{pmatrix} I_{N-1} \\ 0_{1 \times N-1} \end{pmatrix} - \frac{1}{N} \mathbb{1}_{N \times N-1} \\
 & x^f = \sqrt{N} X^f \mathbb{1}_{N \times 1} \\
 & P^a = L^f A^f L^{fT}, \\
 & \text{with } A^f = \frac{1}{N} (T^T T)^{-1} + (L^{fT} L^f)^{-1}, L^{fT} Q L^f (L^{fT} L^f)^{-1} \\
 & P^a = L^f A^a L^{fT}, \\
 & \text{with } A^{a-1} = A^{f-1} + (YT)^T R^{-1} Y T \\
 & x^a = x^f + L^f A^a (YT)^T R^{-1} (y - y^f), \\
 & \text{with } y^f \text{ the observation operator on } x^f
 \end{aligned}$$

The SEIK+PCA Filter

Z is substituted by

$$\tilde{Z} = \sqrt{N} U_r S_r \Omega_r^T,$$

where

$$USV^T = LC$$

is the singular values decomposition of LC .

If r singular values are predominant (with $2r > N - 1$), then $U_r \in \mathbb{R}^{n \times r}$ and $S_r \in \mathbb{R}^{r \times r}$ are the submatrices of U and S that refer to those values.

$\Omega_r \in \mathbb{R}^{r \times N}$ has the same properties as Ω , but it is obtained completing the matrix

$$\begin{pmatrix} I_{N-1-r} \\ -I_{N-1-r} \\ 0_{2r+2-N} \end{pmatrix}.$$

The experiment

The filters has been compared in a one dimensional transport model with 2 biological variables.

One variable provided data in a limited part of the domain (Assimilation type 1: only in the center, assimilation type 2: uniformly distributed points) every 20 timesteps, while the second variable was hidden.

20 different parameter sets has been tried for each case.

The model

$$d_p = Ac_p(M - c_p) - Bc_p c_q,$$

$$d_q = Cc_p c_q - Dc_p,$$

$$\frac{\partial}{\partial t} c_i = -V(x) \frac{\partial}{\partial x} c_i + K \frac{\partial^2}{\partial x^2} c_i, \quad \forall i \in \{p, q\}.$$

Results

	Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10
Type 1	3.22	0.65	1.91	7.90	0.12	0.10	0.95	2.58	1.80	2.41
Type 2	0.80	0.22	0.03	-0.18	0.09	0.26	0.02	1.20	1.47	1.14
	Ex11	Ex12	Ex13	Ex14	Ex15	Ex16	Ex17	Ex18	Ex19	Ex20
Type 1	0.86	7.57	0.86	3.89	-1.14	3.74	0.54	2.46	1.34	0.06
Type 2	1.14	1.51	0.06	-0.09	0.35	0.43	0.09	0.43	-0.29	1.30

Table 1: Difference of root mean square distance between the classic and the modified SEIK filter

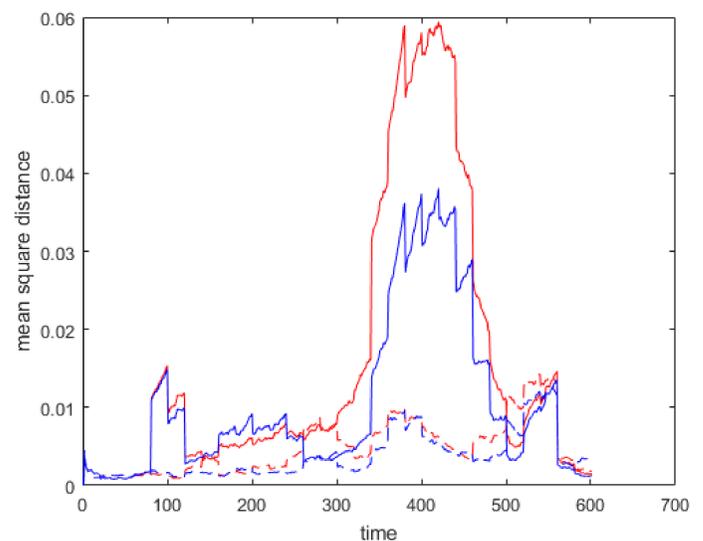


Figure 1: Experiment 10 Type 1. Red: classic SEIK, blue: modified SEIK. The dashed line is the hidden variable, while the other is the assimilated one.

Conclusions

The modified SEIK filter has obtained better results than the classic SEIK filter in 36 of 40 tests made. This confirms that using the PCA to change the error basis when needed helps significantly to achieve better precision, without rising the number of ensemble elements.

Forthcoming Research

Testing the new Kalman scheme on a more complicated model, like the MedBFM System, will be a priority. Implementation and optimization will follow up, if the performances will be confirmed.

In the near future, how the candidates for the substituted basis vectors are chosen is a point that will be better investigated, in order to improve further the skill of the scheme.

References

- [1] *The Biogeochemical Flux Model (BFM), Equation Description and User Manual, BFM version 5.1, Release 1.1*, August 2015.
- [2] G. Cossarini, P. Lazzari, and C. Solidoro. Spatiotemporal variability of alkalinity in the mediterranean sea. *Biogeosciences*, 12(6), 2015.
- [3] S. Dobricic and N. Pinardi. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Modelling*, 2008.
- [4] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288, 2011.
- [5] P. Lazzari, C. Solidoro, V. Ibello, S. Salon, A. Teruzzi, K. Béranger, S. Colella, and A. Crise. Seasonal and inter-annual variability of plankton chlorophyll and primary production in the mediterranean sea: a modelling approach. *Biogeosciences*, 9, 2012.
- [6] P. Lazzari, A. Teruzzi, S. Salon, S. Campagna, C. Calonaci, S. Colella, M. Tonani, and A. Crise. Pre-operational short-term forecasts for the mediterranean sea biogeochemistry. *Ocean Science*, 6, 2010.
- [7] L. Nerger and W. W. Gregg. Assimilation of seawifs data into a global ocean-biogeochemical model using a local seik filter. *J. Mar. Syst.*, 68, 2007.
- [8] A. Teruzzi, S. Dobricic, C. Solidoro, and G. Cossarini. A 3-d variational assimilation scheme in coupled transport-biogeochemical models: Forecast of mediterranean biogeochemical properties. *J. Geophys. Res. Oceans*, 119, 2014.
- [9] G. Triantafyllou, I. Hoteit, and G. Petihakis. A singular evolutive interpolated kalman filter for efficient data assimilation in a 3-d complex physicalbiogeochemical model of the cretan sea. *J. Mar. Syst.*, 41, 2002.