



# Obstacles and benefits of the implementation of a reduced-rank smoother with a high resolution model of the tropical Atlantic Ocean

N. Freychet<sup>1</sup>, E. Cosme<sup>1</sup>, P. Brasseur<sup>2</sup>, J.-M. Brankart<sup>2</sup>, and E. Kpemlie<sup>2</sup>

<sup>1</sup>Université Joseph Fourier – Grenoble 1, LEGI, Grenoble, France

<sup>2</sup>CNRS UMR5519, LEGI, Grenoble, France

Correspondence to: E. Cosme (emmanuel.cosme@legi.grenoble-inpg.fr)

Received: 29 February 2012 – Published in Ocean Sci. Discuss.: 20 March 2012

Revised: 18 August 2012 – Accepted: 21 August 2012 – Published: 18 September 2012

**Abstract.** Most of oceanographic operational centers use three-dimensional data assimilation schemes to produce re-analyses. We investigate here the benefits of a smoother, i.e. a four-dimensional formulation of statistical assimilation. A square-root sequential smoother is implemented with a tropical Atlantic Ocean circulation model. A simple twin experiment is performed to investigate its benefits, compared to its corresponding filter. Despite model's non-linearities and the various approximations used for its implementation, the smoother leads to a better estimation of the ocean state, both on statistical (i.e. mean error level) and dynamical points of view, as expected from linear theory. Smoothed states are more in phase with the dynamics of the reference state, an aspect that is nicely illustrated with the chaotic dynamics of the North Brazil Current rings. We also show that the smoother efficiency is strongly related to the filter configuration. One of the main obstacles to implement the smoother is then to accurately estimate the error covariances of the filter. Considering this, benefits of the smoother are also investigated with a configuration close to situations that can be managed by operational center systems, where covariances matrices are fixed (optimal interpolation). We define here a simplified smoother scheme, called half-fixed basis smoother, that could be implemented with current reanalysis schemes. Its main assumption is to neglect the propagation of the error covariances matrix, what leads to strongly reduce the cost of assimilation. Results illustrate the ability of this smoother to provide a solution more consistent with the dynamics, compared to the filter. The smoother is also able to produce analyses independently of the observation frequency, so the smoothed solution appears more continuous in time, especially in case of a low frequency observation network.

## 1 Introduction

Data assimilation methods for geophysics have evolved continuously since their origins in the 70s. In the branch of estimation theory, the Kalman filter (Kalman, 1960) has been widely used in oceanography. Its implementation with large numerical models is made possible provided relevant adaptations, such as reduced-order formulation (Parrish and Cohn, 1985; Todling and Cohn, 1994; Evensen, 1994; Fukumori and Malanotte-Rizzoli, 1995 and Houtekamer and Mitchell, 1998). The filter provides an optimal estimate of the system state, given a model state (numerical), and past and present observations available up to the analysis time. Thus, it is well indicated to initialize a forecast, which is the historical purpose of data assimilation in geophysics. Though, oceanographic applications of data assimilation get increasingly diversified. In particular, climate studies require accurate reconstructions of the past ocean circulation, *reanalyses*, as performed for instance in the *MyOcean* project (<http://www.myocean.eu.org/>). Such reanalyses are expected to gain accuracy when the observation datasets are used in a four-dimensional way with data assimilation, i.e. when each observation has an influence on past, present and future states of the model solution. In the framework of the Kalman filter, the past influence of observations is introduced through retrospective analyses and leads to a smoothing formulation of the estimation problem. Optimal smoothers in estimation theory may be considered as a four-dimensional extension of the Kalman filter that takes into account future observations.

The relevance of using a smoother for high dimensional oceanic or atmospheric problems is still an open question. Even if linear theory says that smoothing decreases residual

filter errors, the usual approximations (on non-linearity, rank reduction, localisation, etc.) take these problems a long way from theory. In the study of ?, the smoother produced apparently poor improvements over the filter, but the meteorological forecasts started from smoother estimates were better. On the contrary, with very different settings (ocean model, forward-backward smoother), Lermusiaux et al. (2002) obtained better estimates with the smoother (in terms of errors), but poorer forecasts. Khare et al. (2008) tried to identify regimes where the smoother is particularly efficient with an atmospheric model, but this is very case-dependent. The work reported in this paper is part of an effort to determine the relevance of smoothing for realistic oceanic problems.

Several smoother approaches exist (see Cosme et al., 2011, for detailed descriptions). In this paper, we consider the sequential fixed-lag smoother (Cohn et al., 1994; Evensen and van Leeuwen, 2000), well designed for reanalysis purposes. More precisely, we use the reduced-rank square-root smoother developed by Cosme et al. (2010), which is based on the singular evolutive extended Kalman (SEEK) filter (Pham et al., 1998; Brasseur and Verron, 2006). Actually, the *MyOcean* reanalysis system is based on the SEEK filter too. This reduced-rank smoother has been tested by Cosme et al. (2010) with a square-box configuration of a high resolution ocean circulation model. In this work, we investigate the application of the reduced-rank smoother with a more complex and realistic tropical Atlantic Ocean circulation model in a  $1/4^\circ$  resolution configuration. We strive to identify obstacles and solutions to implement the smoother in such a realistic context. The gain of the smoother over the filter is assessed. Finally, to comply with operational constraints, we design here a new flavour of the sequential fixed-lag smoother algorithm, referred to as *half-fixed basis*, to overcome implementation issues. Our main goal is to determine whether this smoother can improve present reanalysis schemes.

In Sect. 2, we present the SEEK filter and the SEEK smoother formulations as detailed in Cosme et al. (2010). Some advantages and drawbacks linked to implementation are stressed. Section 3 summarizes the set-up of our twin experiments. The smoother is then implemented according to its theoretical formulation. In Sect. 4 we dwell upon a sensitive step of the implementation of a smoother: the parametrisation and the dynamical propagation of error covariances. In Sect. 5 we expose and examine the main results of a short smoother reanalysis. The filter reanalysis is used as a reference to point out the improvements of a four-dimensional extension of the data assimilation. To deal with implementation issues raised in Sect. 4, a half-fixed basis smoother algorithm is developed and tested in Sect. 6. Section 7 concludes and gives perspectives.

## 2 The reduced-rank square-root filter and smoother algorithms

The Kalman filter and smoother formulations can be found in Anderson and Moore (1979), Simon (2006) and Evensen (2007). For the smoothers, see also Cosme et al. (2011). Here we only provide an overview of the sequential algorithms, close to the EnKS, and an intuitive interpretation of the equations. We also delineate these algorithms in their square-root transformation so they can be implemented in large geophysical problems. Finally, we expose the modifications introduced for the half-fixed basis smoother to be implemented.

### 2.1 Notations

Notations are similar to the ones used in Cosme et al. (2010). Superscripts refer to the type of states (“a” for analysed and “f” for forecast). Subscripts are used to specify the transition between two times (for instance  $k-1, k$  is used for the transition matrix between times  $k-1$  and  $k$ ) or the time of a state and information it contains (for instance  $k|k-1$  indicates that the state is estimated at time  $t_k$ , given all observations available up to time  $t_{k-1}$ ). Subscripts 0 represent the initial conditions.

### 2.2 The Kalman filter

The sequential form of the filter is given by a succession of two steps, forecast and analysis, summarized in Eqs. (1)–(7). Initial conditions must first be prescribed:  $\mathbf{x}_0^a$ , an initial state, and  $\mathbf{P}_0^a$ , an initial error covariance matrix.

Initialization  $\mathbf{x}_0^a$  and  $\mathbf{P}_0^a$

Forecast step

$$\mathbf{x}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1|k-1}^a \quad (1)$$

$$\mathbf{P}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{P}_{k-1|k-1}^a \mathbf{M}_{k-1,k}^T + \mathbf{Q}_{k-1,k} \quad (2)$$

Analysis step

$$\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_{k|k-1}^f \quad (3)$$

$$\mathbf{G}_k = \mathbf{H}_k (\mathbf{P}_{k|k-1}^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (4)$$

$$\mathbf{K}_{k|k} = (\mathbf{P}_{k|k-1}^f \mathbf{H}_k^T) \mathbf{G}_k^{-1} \quad (5)$$

$$\mathbf{x}_{k|k}^a = \mathbf{x}_{k|k-1}^f + \mathbf{K}_{k|k} \mathbf{d}_k \quad (6)$$

$$\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathbf{K}_{k|k} \mathbf{H}_k) \mathbf{P}_{k|k-1}^f \quad (7)$$

The filter performs a forecast step by propagating  $\mathbf{x}_{k-1|k-1}^a$  with the dynamical linear model operator  $\mathbf{M}_{k-1,k}$  (Eq. 1), leading to the forecast state  $\mathbf{x}_{k|k-1}^f$  at time  $t_k$ . Error covariance matrix  $\mathbf{P}_{k-1|k-1}^a$  is also propagated with the model operator (Eq. 2), and a model error  $\mathbf{Q}_{k-1,k}$  is added to take into

account model uncertainties and approximations. Thus, at the end of the forecast step (at time  $t_k$ ), an estimate of the ocean state  $\mathbf{x}_{k|k-1}^f$  and error statistics  $\mathbf{P}_{k|k-1}^f$  associated to this state are provided by the filter, given all observations available up to time  $t_{k-1}$ .

Observations at time  $t_k$  are then used to perform the filter analysis step. First the innovation vector  $\mathbf{d}_k$  is computed (Eq. 3), providing the difference between observations ( $\mathbf{y}_k$ ) and the model state ( $\mathbf{x}_{k|k-1}^f$ ) projected into the observation space with the observation operator  $\mathbf{H}_k$ . The innovation error covariance matrix  $\mathbf{G}_k$  is also defined (Eq. 4), with  $\mathbf{R}_k$  the observation error matrix.  $\mathbf{G}_k$  is then used to compute the Kalman gain  $\mathbf{K}_{k|k}$  (Eq. 5). Then, the update of the forecast state,  $\mathbf{x}_{k|k}^a$ , is computed (Eq. 6), balancing observations and the model estimate thanks to the Kalman gain. Statistics on the residual error,  $\mathbf{P}_{k|k}^a$ , are also estimated through Eq. (7). At the end of the analysis step, the Kalman filter provides the best state estimate given all observations available up to time  $t_k$ .

Analysis and forecast steps are performed successively from the first to the last observed date.

### 2.3 The sequential smoother

The smoother uses observations at a time  $t_k$  to improve past estimates at times  $t_i$ , with  $t_i < t_k$ . Thus, an analysed state at time  $t_i$  is now noted  $\mathbf{x}_{i|k}^a$ , meaning it contains all informations from observations available until time  $t_k$ . Obviously,  $i$  is not unique, meaning observations at time  $t_k$  can be used by the smoother to analyse several past states (at different times  $t_i$ ). The set of time indices at which the retrospective analyses are produced is noted  $\Sigma_k$ . In our configuration,  $\Sigma_i = \{k-L, \dots, k-1\}$ , corresponding to a fixed-lag smoother according to the nomenclature used in Cosme et al. (2010). With this configuration, the smoother performs retrospective analysis for the  $L$  states previous to time  $t_k$  only ( $L$  being called the *lag* of the smoother). This can be interpreted as *time localisation* of the smoother, restricting the past influence of observations. It also limits the numerical storage of the smoother.

Forecast step

$$\mathbf{x}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1|k-1}^a \quad (8)$$

$$\mathbf{P}_{k,i|k-1}^{fa} = \mathbf{M}_{k-1,k} \mathbf{P}_{k-1,i|k-1}^{aa}, \quad i \in \Sigma_k \quad (9)$$

$$\mathbf{P}_{k|k-1}^f = \mathbf{M}_{k-1,k} (\mathbf{P}_{k,k-1|k-1}^{fa})^T + \mathbf{Q}_{k-1|k} \quad (10)$$

Analysis step

$$\mathbf{K}_{i|k} = \mathbf{P}_{k,i|k-1}^{fa} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1}^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (11)$$

$$\mathbf{x}_{i|k}^a = \mathbf{x}_{i|k-1}^a + \mathbf{K}_{i|k} \mathbf{d}_k, \quad i \in \Sigma_k \quad (12)$$

$$\mathbf{P}_{k,i|k}^{aa} = (\mathbf{I} - \mathbf{K}_{k|k} \mathbf{H}_k) \mathbf{P}_{k,i|k-1}^{fa}, \quad i \in \Sigma_k \quad (13)$$

$$\mathbf{P}_{i|k}^a = \mathbf{P}_{i|k-1}^a - \mathbf{K}_{i|k} \mathbf{H}_k \mathbf{P}_{k,i|k-1}^{fa}, \quad i \in \Sigma_k \quad (14)$$

Equations of the optimal linear fixed-lag smoother are summarized in Eqs. (8)–(14). To perform the retrospective analysis, the smoother needs the introduction of cross-covariances matrices, that is, matrices of covariances between state errors at two different times. These matrices are involved in the smoother gain (Eq. 11), which is used for the smoother analysis (Eq. 12). Thereby, the cross-covariance matrices enable the use of information at time  $t_k$  to correct a past state at time  $t_i$ . Note that in the forecast step, the Eq. (2) of the filter is split into two equations for the smoother (Eqs. 9 and 10) to bring out the cross-covariances matrix.

### 2.4 The square-root transformation of the Kalman filter and smoother

We now expose algorithms in a reduced-rank form so they can be applied to large geophysics systems. We use the singular evolutive extended Kalman (SEEK) filter that has already been implemented with real systems (e.g. Verron et al., 1999; Testut et al., 2003; Brankart et al., 2003 and Castuccio et al., 2006). A synthesis on the SEEK filter can be found in Brasseur and Verron (2006) or Rozier et al. (2007). The main idea is to use a square-root decomposition of the error covariance matrix so it can be written as:  $\mathbf{P}^f = \mathbf{S}^f \mathbf{S}^{fT}$ , where  $\mathbf{S}^f$  is a  $n \times n$  matrix,  $n$  being the length of the state vector. The filter equations are then reformulated including the square-root decomposition, as exposed in Eqs. (15)–(24).

To be computable at low cost, some assumptions are introduced. First, the dimension of  $\mathbf{S}^f$  can be reduced assuming that errors only occur on a low-dimensional subspace,  $r$ , of the state space (with  $r \ll n$ ). In practice, these error directions can be identified keeping the  $r$  first empirical orthogonal functions (EOFs) computed from a time series of model states. When considering only the  $r$  main directions of the error,  $\mathbf{S}^f$  becomes a  $n \times r$  matrix, the columns of which are often referred to as *error modes*. The term  $\mathbf{I} + \Gamma$  in the smoother gain is then  $r \times r$ , and easily invertible if  $r$  is small enough. Moreover, the propagation of error covariances (Eq. 16) is more affordable (only  $r$  model iterations are needed instead of  $n$ ). The model error term is now noted  $\delta$  in this equation. Another assumption is made about  $\mathbf{R}$ , so it can be easily invertible in the Kalman gain equation: it is considered diagonal here. Note that other assumptions can be used (as presented in Testut et al., 2003; Brankart et al., 2003, 2009) to introduce observation error correlations while keeping the matrix diagonal.

Equations of the SEEK filter and smoother.

Initialization  $\mathbf{x}_0^a$  and  $\mathbf{S}_0^a$

Forecast step

$$\mathbf{x}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{x}_{k-1|k-1}^a \quad (15)$$

$$\mathbf{S}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a + \delta_{k-1|k} \quad (16)$$

Filter analysis step

$$\Gamma_k = (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f) \quad (17)$$

$$\mathbf{K}_{k+1} = \mathbf{S}_{k|k-1}^f [\mathbf{I} + \Gamma_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (18)$$

$$\mathbf{d}_{k+1} = \mathbf{y}_{k+1} - \mathbf{H}_{k+1} \mathbf{x}_{k|k-1}^f \quad (19)$$

$$\mathbf{x}_{k|k}^a = \mathbf{x}_{k|k-1}^f + \mathbf{K}_{k|k} \mathbf{d}_k \quad (20)$$

$$\mathbf{S}_{k|k}^a = \mathbf{S}_{k|k-1}^f (\mathbf{I} + \Gamma_k)^{-1/2} \quad (21)$$

Smoother analysis step

$$\mathbf{K}_{i|k} = \mathbf{S}_{i|k-1}^a [\mathbf{I} + \Gamma_k]^{-1} (\mathbf{H}_k \mathbf{S}_{k|k-1}^f)^T \mathbf{R}_k^{-1} \quad (22)$$

$$\mathbf{x}_{i|k}^a = \mathbf{x}_{i|k-1}^a + \mathbf{K}_{i|k} \mathbf{d}_k, \quad i \in \Sigma_k \quad (23)$$

$$\mathbf{S}_{i|k}^a = \mathbf{S}_{i|k-1}^a [\mathbf{I} + \Gamma_k]^{-1/2}, \quad i \in \Sigma_k \quad (24)$$

With the SEEK formulation of the filter, the sequential smoother implementation becomes straightforward. Only three extra equations are needed (Eqs. 22, 23 and 24). The cross-covariance terms are here directly introduced in the smoother gain. The smoother analysis is performed using the smoother gain and the innovation vector computed from the filter. Finally, the smoother analysis covariances are computed. The smoother implementation does not require additional assumption nor significant extra CPU times, with respect to the filter. The only limitation lies in the storage of the smoother covariance matrices, but with the fixed-lag formulation, the number of retrospective analyses, and thus the number of covariance matrices to store for each observation, is limited. The smoother also presents the advantage of being applicable simultaneously to or after the filter (in this case, calculations of the gains must be performed again).

## 2.5 Model error and evolutive covariances

A smoother requires accurate cross-covariances between the filter forecast state and past estimates (Eq. 22). Thus, the filter must provide accurate forecast (and analysis, for the next steps) error modes.

Theoretically speaking, the forecast error modes result from the combination of the dynamical propagation of past analysis modes, which ensures the statistical connection through time, and a white-in-time model error, which makes the short-term cross-temporal decorrelation of modes occur (Eq. 16). Hence, both elements (dynamical propagation and model error) play a key role in the behaviour of the smoother. In particular, it is theoretically obvious that a smoother must

rely on an *evolutive* filter (with dynamical propagation of the modes), termed in this way in contrast to a *static* filter such as ensemble optimal interpolation (Evensen, 2003). It is self-evident that the model error must be accounted for accurately.

The model error parameterisation in ensemble filters has long been and still remains a considerable issue, and is still an active topic of research (Houtekamer et al., 2009; Brankart et al., 2010). The *covariance inflation* approach (Li et al., 2009) is quite popular but inadequate to decorrelate error modes through time (Cosme et al., 2010). Recent efforts have been undertaken for a better, adaptive estimation of the forecast error covariances in the SEEK filter (Brankart et al., 2010, 2011), but further work is necessary to make them robust and applicable routinely. In order to disconnect the smoother and the model error issues here, we have decided to stick to a *perfect model* set-up and neglect the model error term  $\delta$ . This obviously puts limitations on the scope of our experiments, as detailed in Sect. 4. This assumption leads to a new form of Eq. (16):

$$\mathbf{S}_{k|k-1}^f = \mathbf{M}_{k-1,k} \mathbf{S}_{k-1|k-1}^a. \quad (25)$$

The propagative term  $\mathbf{M}\mathbf{S}^a$  implies  $r$  model iterations,  $r$  being the number of columns in  $\mathbf{S}^a$ . To be affordable at a reasonable cost,  $r$  must be kept quite small. A remaining problem lies in the divergence of the filter that can occur due to the order reduction, i.e. the error estimated by the filter ( $\mathbf{S}^f$  and  $\mathbf{S}^a$ ), and can become inconsistent with the true error. The estimation of the errors through time can be controlled thanks to several parameters. This point will be discussed in Sect. 4 with the implementation of an evolutive filter.

Finally, note that even if  $\delta$  is neglected in the covariance propagation, retrospective influence of observations is artificially limited thanks to the fixed-lag form of the smoother.

## 2.6 A half-fixed basis smoother

The model error parameterisation issue, combined with the prohibitive cost of the modes propagation, often leads to use static filters, as various oceanographic operational centers have already turned to (Brosseur et al., 2005; Martin et al., 2007; Oke et al., 2008). It can be expected that these operational centers will have to adapt their assimilation schemes to perform the future reanalyses of the ocean circulation, as has already been done (Ferry et al., 2010), and as is the general case now for the atmospheric circulation. For this reason, and in spite of the theoretical concerns mentioned previously, it seems worth trying to design a smoother based on a static filter (minimizing the impacts of the violations in the theory) and testing it numerically.

In this configuration, only Eq. (25) needs to be rewritten in a static form:  $\mathbf{S}_{k|k-1}^f = \mathbf{S}_0^a$ . Kalman gains of the filter (Eq. 18) and the smoother (Eq. 22) are also affected, with the terms  $\mathbf{S}_{k|k-1}^f$  being replaced by  $\mathbf{S}_0^a$ . Another term in Eq. (22) needs to be specified: the analysis error covariance matrix  $\mathbf{S}_{i|k-1}^a$ .

This matrix can be considered as fixed (so the smoother would also be a fixed basis type), but a more theoretically sound approach consists of using the analysis covariance matrix computed from  $\mathbf{S}_0^a$  with the filter at the appropriate time. This smoother scheme is referred to as half-fixed basis, because a part of the cross-covariance matrix is kept fixed ( $\mathbf{S}_{k|k-1}^f$ ) while the other part is time dependent ( $\mathbf{S}_{i|k-1}^a$ ). We will expose and discuss more in detail this smoother in Sect. 6.

### 3 Twin experiments and assimilation settings

The smoother is implemented and assessed in the framework of twin experiments, but with the concern of remaining close to a real problem (through model and simulated observations). The experimental settings are described in this section. The nomenclature of experiments is summarized in Table 2.

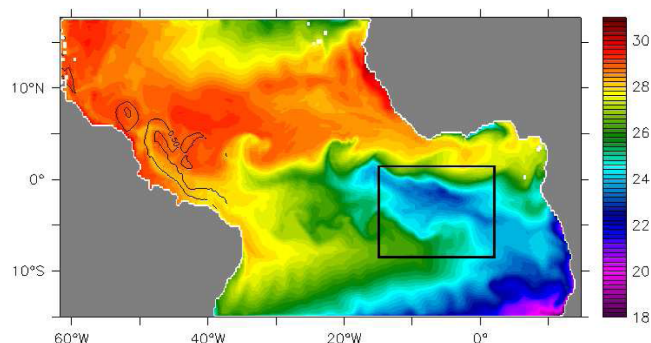
#### 3.1 Model and configuration

The ocean circulation model NEMO (Nucleus for European Modelling of Ocean, Madec, 2008) is used in a tropical Atlantic configuration (from 61.5° W to 15° E, and 15° S to 17.75° N) and with a resolution of 1/4 degree. This configuration, previously developed and used by Ubelmann et al. (2009), is named TATL4 (Tropical ATLantic 1/4 degree). Boundary conditions are extracted from a model simulation spanning the 1958–2007 period (Dussin et al., 2009). An initial condition for TATL4 is extracted on 5 January 1995. Atmospheric forcings (Brodeau et al., 2010) are identical. An interannual simulation is then performed until the end of 2005. This run is used as a reference for the twin experiment and is called REF.

Two main signals dominate the dynamics of the tropical Atlantic, as illustrated by Fig. 1. On the one hand, tropical instability waves (TIWs) develop during summer and fall at 3° N. They propagate from east to west and exhibit a strong signal on the sea surface temperature (Legeckis, 1977; Allen et al., 1995). On the other hand, eddies develop and circulate along the north-east coast of Brazil. These eddies, known as Brazil rings (Richardson and Walsh, 1986; Garzoli and Katz, 1993), are due to the North Brazil Current retroflexion and exhibit a strong signal on velocity (from the surface to 200–500 m depth). Due to their propagative signatures, these two signals are interesting objects of study for a four-dimensional assimilation experiment.

#### 3.2 Perturbation

To perform a twin experiment, a perturbed ocean state is generated using the interannual variability. An initial condition from REF on 25 May 2003 is used to restart the model on 25 May 2005. This error on initial condition leads to a perturbed simulation called FALSE, computed until the end of



**Fig. 1.** TATL4 configuration: sea surface temperature is colored and exhibits the development of TIWs in the center of the basin; black contours indicate the 0.5 m s<sup>-1</sup> iso-velocity at 30 m depth, exhibiting the North Brazil Current and the rings formation. The black rectangular box illustrates the influence zone of an observation (see Sect. 3.5).

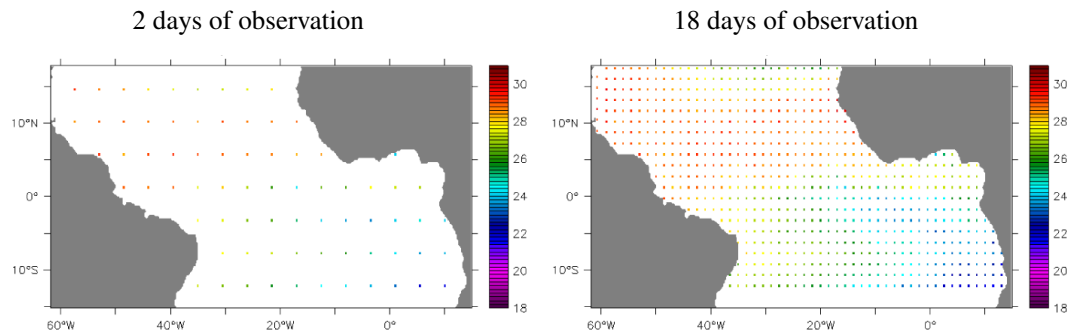
2005. The data assimilation experiment starts one month after the beginning of FALSE, so that the perturbed simulation can partially adjust with the forcings and damp, spurious, high-frequency oscillations. The goal of the present data assimilation experiment is then to correct FALSE so it becomes closer to REF. Note that even if a perturbation of the initial condition is not sufficient to assess the performance of the smoother for a long-term reanalysis, it allows the careful study of the impact of the smoother analysis and the quality of corrected states, at least at the beginning of the experiment.

#### 3.3 Simulated observations

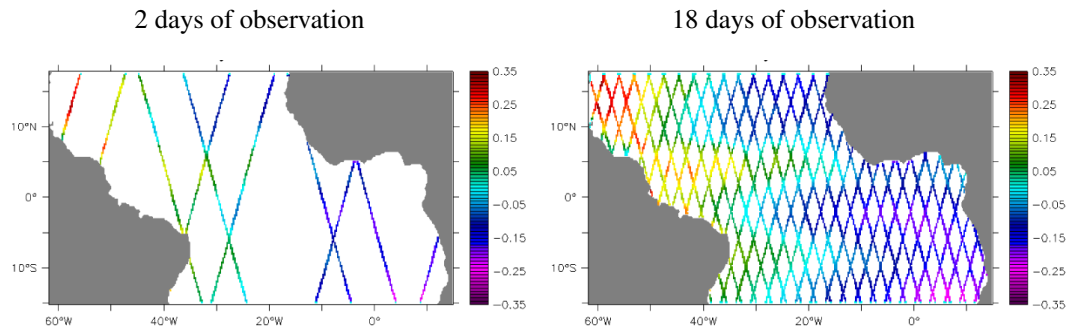
Synthetic observations are extracted from REF following the procedure described by Cosme et al. (2010). Two sets of observations are generated: in situ data of temperature and salinity, and altimetric measurements.

In situ data mimic ARGO profilers network. Every two days, a set of vertical profiles is available, with the profiles 6° apart from each other. This pattern is shifted by 2° from one assimilation step to the next. It results in a 2° density network every 18 days, close to the average true ARGO network (one profile every 2° and every 15 days). The simulated in situ data network is illustrated on Fig. 2, for one assimilation cycle (every 2 days) and the full coverage (after 18 days).

Altimetric data mimic Jason-type satellite tracks. Every 2 days, a limited number of tracks of sea surface height (SSH) are available. The periodicity of these tracks is 10 days. Figure 3 provides an example of SSH tracks available for one assimilation cycle (every 2 days) and shows the complete coverage obtained after 10 days of consecutive altimetric observations.



**Fig. 2.** Horizontal distributions of in situ observation network simulated for the experiment. Left: observations available for one assimilation step (here: 30 June 2005). Right: observations coverage after 18 days.



**Fig. 3.** Horizontal distributions of altimetric observation network simulated for the experiment. Left: observations available for one assimilation step (here: 30 June 2005). Right: observations coverage after 10 days.

We consider the measurement error to be null here: the observations are not perturbed and can be described as *perfect observations*. However, the matrix  $\mathbf{R}$  is not zero because of the truncation error. Its parametrisation is presented in Sect. 4. The preliminary experiments have been carried out this way, and since the appropriate value of  $R$  to account for representativeness is much larger than the actual observation error, perturbing the observations was considered somewhat superfluous. Re-running the experiments would be extremely expensive and would not change the results significantly.

### 3.4 Initial statistics

As the only source in the initial state error is connected to the model's time variations, the statistics  $\mathbf{S}_0^a$  describing the initial error covariance are calculated based on the model's time variability. 270 EOFs are computed from REF, using ocean states in summer and fall between 1995 and 2000. Winter and spring are not considered because the experiment spans the summer months only, and because TIW or Brazil ring signatures are mostly present in summer and fall. The first 39 EOFs, which represent about 95 % of the total variance of the 270 initial EOFs, are selected to form the initial error modes. Note that in an operational forecast system, where additional errors (in forcings for instance) have to be considered, the number of degrees of freedom for the errors would

be considerably increased, and hence the required size of the ensemble.

### 3.5 Localisation

A consequence of the order reduction is that statistics may not be able to faithfully describe all the correlations, especially long-distance correlations (Houtekamer and Mitchell, 1998). To prevent the spurious influence of distant observations during an analysis step, the filter is used with a localisation scheme (Brankart et al., 2009). This method rules out long-distance corrections, so that only observations in the neighbourhood of an analysed point are used for the correction. The size of the neighbourhood is defined as  $15^\circ$  zonal and  $10^\circ$  meridional in length (illustrated by the black box in Fig. 1), a trade-off between analysis accuracy and computational efficiency made after a large number of sensitivity experiments. It is large enough so that at least one observation is available for each analysis. The zonal length is larger than the meridional length because the dynamics in the tropical region are mostly zonal, so the correlations are more consistent in this direction.



#### 4 Evolutive filter

We now present the results of the SEEK smoother in its original form, i.e. with evolving error modes, as presented in Cosme et al. (2010). For the reasons mentioned in Sect. 2.5, we pay particular attention to the accurate estimation of the forecast error modes (defined by Eq. 25 in our configuration). It is well known that due to the truncature of **S** essentially, the error modes can collapse or become inconsistent with the true error (error on the mean) with time. The most common method to avoid this is to use covariance inflation. However, for the reasons given in Sect. 2.5, it is not our choice here. Instead, we chose to tune a few parameters controlable in the system: the number of error modes (columns in **S**), the size of the localisation domain, and the amplitude of the (diagonal) observation error matrix **R**. The adopted criterion of consistency between errors on the mean and errors estimated by the modes is that the RMS error computed with the mean,

$$\text{RMS} = \sqrt{\frac{1}{N \times M \times P} \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^P (x_{i,j,k} - x_{i,j,k}^{\text{ref}})^2}$$

(with  $i = 1 : N$  the set of the longitudes,  $j = 1 : M$  the set of the latitudes,  $k = 1 : P$  the set of the vertical levels,  $x_{i,j,k}$  the mean state of the ocean and  $x_{i,j,k}^{\text{ref}}$  the true state of the ocean, REF), must be of the same order of magnitude as the square root of the trace of the covariances matrix ( $\text{tr}(\mathbf{SS}^T)$ ).

The first two parameters, though, are partly constrained by practical considerations. A high number of error modes allows an accurate representation of the error space, but requires more CPU time for modes propagation. Several sensitivity experiments, not detailed here, suggested that 39 is a fair number. The localisation parameters must also be prescribed optimally to avoid spurious corrections due to distant observations and no correction due to absence of close observations. Again, sensitivity experiments showed that having at least one observation for each analysed grid point was a fair choice.

The remaining parameter to be specified is the matrix **R**, which represents the confidence of the observations and of the resulting analysis. This matrix includes the observation error (including the accuracy of measurements), but not only that. As mentioned in Sect. 2.4, this matrix is considered diagonal. This assumption implies that the diagonal terms should be inflated to include the ignored non-diagonal terms. Moreover, the truncation error (due to the reduction order) should also be included in **R**. This error, as well as the non-diagonal terms of **R**, are not easy to quantify, and a wrong definition of the diagonal terms of **R** can lead to inconsistent error statistics with the real error. Precedent studies point out these aspects, as for instance Liu and Rabier (2002), Rabier (2006) and Oke et al. (2008). To correctly parametrise **R**, several filtering experiments are conducted with different values for observation errors, as specified in Table 1. The experiments lasted 50 days, starting from 25 June, considered as

**Table 1.** Observation errors of SSH, temperature and salinity for sensitivity experiments ErrObs01 to ErrObs04.

Experiment name	SSH (cm)	Temperature (°C)	Salinity
ErrObs01	0.03	0.8	0.5
ErrObs02	0.08	1.5	1
ErrObs03	0.15	2	1.5
ErrObs04	0.3	3	3

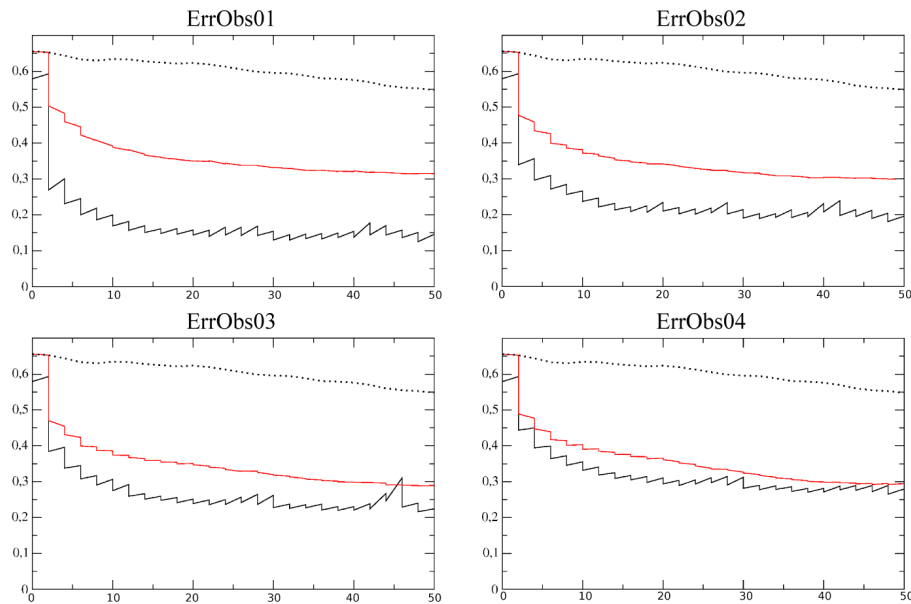
day 0 for the results. For each experiment, the true RMS error is compared to the mean estimate error of the filter ( $\text{tr}(\mathbf{SS}^T)$ ). Results are illustrated with the temperature error levels on Fig. 4. The red curve represents the true error, and the full black curve the estimated error. The error of FALSE is also shown with the black dotted line. To be consistent, the estimated error (black) should be close to the true error (red).

Results show that for the error statistics to be consistent, prescribed observation errors have to be large. Indeed, the best estimated error is obtained with the largest observations error (ErrObs04 in Table 1). This suggests that the truncature error included in **R** is ten times larger than a traditional measurement error. On the other hand, the mean true error (in red) is poorly influenced by this parametrisation. Consequently, large observation errors avoid the collapse of error modes without affecting the quality of the mean state estimate. Also, as a (negative) counterpart of neglecting the model error, the filter converges to an asymptotic level of error, which makes the smoother assessment relevant in the first few weeks of the experiment only. The results are similar for all other assimilated variables (salinity and SSH, not shown). The filter is less efficient for velocities, but still the estimated error on these variables is better with high observation errors, as for the assimilated variables.

The parametrisation of an evolutive filter is difficult. The definition of the error covariance matrix and the control of its time evolution is tricky, and may lead to consideration of unusual values for certain parameters (such as the observations error in our case). This would be especially true with a more complex system where error sources would not only be linked to the initial condition, and it certainly is a major obstacle for the smoother implementation. A strategy to avoid this problem will be exposed in Sect. 6, but first we discuss the smoother effects with the previously defined evolutive filter in the next section.

#### 5 Smoother’s impact

The smoother is tuned using the parameters defined for the SEEK filter in the previous section. The fixed-lag parameter, i.e. the temporal localisation of influence of the observations, is set to 10 days (5 retrospective assimilations). Extending the lag to more than 10 days barely improves the smoother results. This is due to the decrease in amplitude of



**Fig. 4.** Influence of temperature observation error on the error statistics. Red curve represents the true RMS error on temperature and black curve the error estimated with the filter (y-axis: day of the experiment, x-axis: RMS error, in °C). The black dotted curve shows the RMS error of FALSE. Each figure corresponds to a different parametrisation of temperature error in  $\mathbf{R}$ . Values are specified in Table 1 .

the cross-covariance matrix  $\mathbf{S}^a \mathbf{S}^f T$  with time, either because  $\mathbf{S}^a$  and  $\mathbf{S}^f T$  become orthogonal due to the model dynamics working on the latter, or because of the continuous decrease of the former due to the successive retrospective analyses.

### 5.1 Smoothing improvements on error statistics

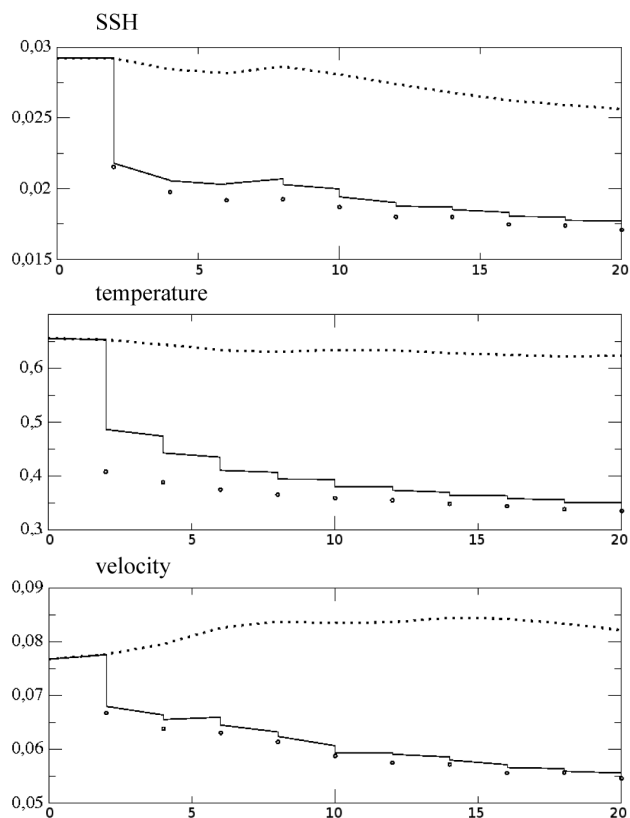
The smoother impact on the mean error levels is shown in Fig. 5 for two assimilated variables (SSH and temperature) and a non-assimilated variable (zonal velocity). As expected, the error on SSH or velocity is reduced with the smoother, though to a limited extent, due to relatively small initial errors in SSH. In the tropics, where the large scale dynamics are not dominated by geostrophy, variations in SSH are low in amplitude, and therefore hard to control with data assimilation (Ubelmann et al., 2009). The small improvements in velocity are due to the fact that corrections on this variable mostly occur near the surface, and errors are computed over the 3 spatial dimensions. The signature of the smoother impact is then mitigated by the 3-D signal. Results are more contrasted for the thermodynamical variables: temperature and salinity. Compared to the filter analysis, the mean error level of the smoothed states is 10 % to 15 % lower at the beginning of the experiment (before day 10). After day 10, the smoother effects tend to decrease, concomitantly with the filter corrections. This behaviour is satisfactory, and directly related to the settings of our experiment, where the errors are introduced in the initial state only. In such settings, the filter converges with time toward an asymptotic error level, and the innovations tend to be zero. Like the filter correction,

the smoother corrections are based on the filter innovation (Eq. 19). If the innovation is null, whatever the reason is, the smoother will not bring any correction. A smoother is not meant to counterbalance spurious tuning of the filter, but only to make optimal (4-D) use of observations, especially when they are sparse in nature, space, and time. In the case of a reanalysis, the filter is expected to have a significant impact during all the experiment, along with the smoother. As we are interested in the smoother effects here, we now focus on the first part of the experiment when the smoothed error level is significantly lower than the filtered error level.

### 5.2 The smoother's dynamics estimation

The smoother is able to reduce the error level of the filter analysis. We now examine its capacity to estimate dynamically balanced states of the ocean circulation. Indeed, a Kalman filter analysis estimate results from a trade-off between dynamics (from the background estimate) and statistics (from the correction term). However, it is still supposed to represent the state of a dynamical system and, as such, must be *dynamically balanced*, that is, respectful of the dynamics represented by the system's model. A smoother estimate must be, at least in theory, closer to the truth than the filter estimate. However, it has also undergone a series of statistically-based corrections that move this dynamics/statistics trade-off toward statistics, which may further alter its dynamical balance. To check the representation of the dynamics in the smoothed states (compared to the filtered states), we perform a simple test: analysed states (from the filter or the smoother at day 2, i.e. the first analysed date)

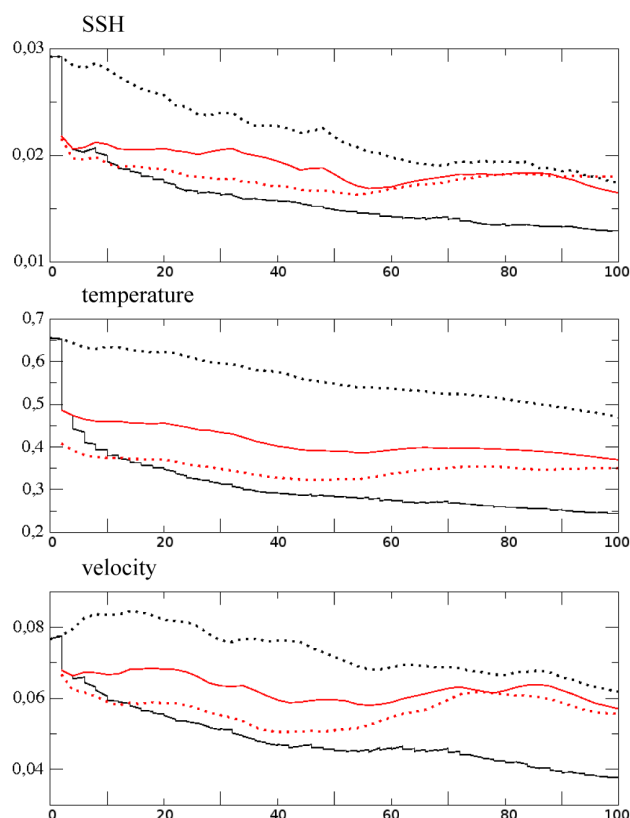




**Fig. 5.** RMS errors for the SSH (m), the temperature ( $^{\circ}\text{C}$ ) and the zonal velocity ( $\text{m s}^{-1}$ ), for the filter (full black line) and the smoother (circles). The trajectories of the run FALSE are also represented by a dotted line.

are used to initialize a free run of the model (respectively FreeFilter and FreeSmooth in Table 2). The implicit idea is that if a state is dynamically inconsistent, this should be reflected in the error evolution, as illustrated by Rozier et al. (2007).

Results for error levels in SSH, temperature and zonal velocity are shown on Fig. 6. First, the error level at day 10 is the same for the filtered run (full black line) and the FreeSmooth run (dotted red line), for all variables, and this is a notable point. Indeed, the same observations are used in both cases, but the filter assimilates them sequentially in time while the smoother assimilates them on the same date (here, day 2). It means that the smoother could extract the same information from observations, but corrected the same state. More generally, the evolution of error levels (on both variables) for FreeSmooth is better during the first 50 days. Beyond, errors tend to converge to the same level (though more slowly for the temperature), becoming close to the FALSE error level. These results suggest that the smoother corrections are consistent with the dynamics, i.e. the reduction of the error level has a real significance from a dynamical point of view and does not introduce significant instabilities.

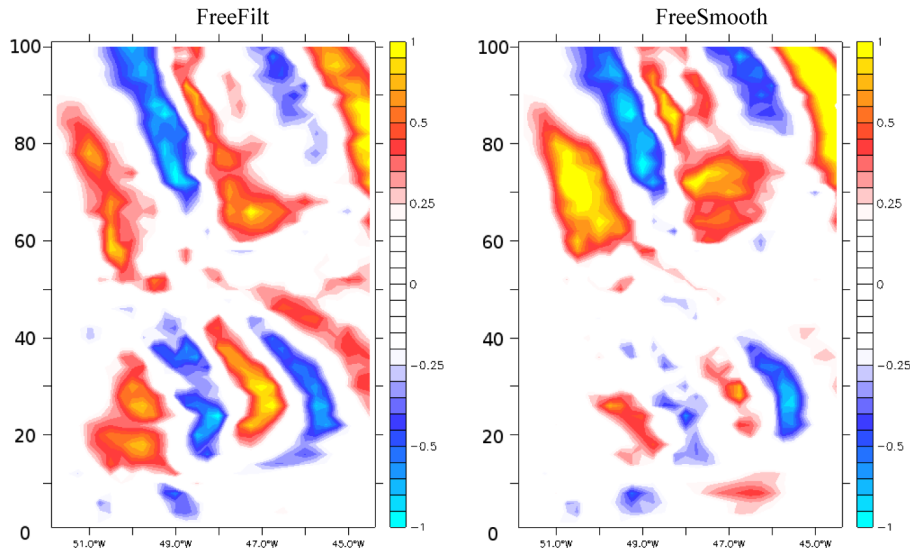


**Fig. 6.** RMS errors for the SSH (m), the temperature ( $^{\circ}\text{C}$ ) and the zonal velocity ( $\text{m s}^{-1}$ ), for the filtered run (full black line) and the free runs restarted with the filtered state at day 2 (full red line) or the smoothed state at day 2 (dotted red line). The RMS errors for FALSE are also shown with a dotted black line.

**Table 2.** Nomenclature of the different runs.

Run name	Description
REF	Interannual run from 1995 to 2005, used as the reference.
FALSE	Perturbed run, initialized on 25 June 2005 with the state of REF from 25 June 2003.
FreeFilter	Run initialized on day 2 of the data assimilation experiment, with the filtered state.
FreeSmooth	Run initialized on day 2 of the data assimilation experiment, with the smoothed state.
FreeOI	Run initialized on day 10 of the data assimilation experiment, with the state analysed with the optimal interpolation.
FreeSmoothOI	Run initialized on day 10 of the data assimilation experiment, with the smoothed state basis on optimal interpolation.

A region particularly well-indicated to illustrate this point is the North Brazil Current zone. Because of its chaotic dynamics, the initial condition strongly affects the subsequent circulation in this region (especially on the Brazil rings).



**Fig. 7.** Evolution of absolute error on velocity (in  $\text{m s}^{-1}$ ), between REF and FreeFilt (left) and FreeSmooth (right). The error is calculated along a section localized at  $5^\circ$  N, and between  $40^\circ$  W and  $50^\circ$  W.

Figure 7 shows the evolution of the error (in velocity) between the free runs and the reference along a transect (between  $44^\circ$  W and  $52^\circ$  W) at  $5^\circ$  N. Figure 8 shows snapshots of the absolute velocity around this same transect, for different times. The alternation between positive and negative anomalies (Fig. 7) is typical of phase shifts affecting the propagative structures, here the Brazil rings. Until day 50, this phase shift is stronger in FreeFilt. If the rings seem quite similar in the filtered state and the smoothed state at day 0 (Fig. 8), they clearly show different evolutions. Rings in FreeSmooth are much closer to the rings in REF. For instance, a ring is correctly formed at day 20 or 30 for FreeSmooth, but this same ring is not completely untied from the North Brazil Current with FreeFilt. Thus, even if not directly visible in the analysed states (day 0), the dynamics are obviously more coherent with REF in the smoothed state. Still, error levels on the free runs (Fig. 7) and snapshots (Fig. 8) after day 50 show that the dynamics in FreeSmooth tends to become also unphased with REF (though the dynamics are not exactly the same as in FreeFilt).

Given all these considerations, it is obvious that the smoothed solution leads to a better estimation of the dynamics, compare to a filtered solution at the same date. The four-dimensional use of the information from the observations has a real impact on analysed states. If the parametrisation linked to the filter is optimal, then the implementation of a smoother gives a better estimated solution, from both statistical and dynamical points of view. In our experiment though, these results are limited to the beginning of the experiment due to the settings of the problem.

Despite its benefits, a remaining obstacle in implementing a smoother is the parametrisation of the evolutive filter. Even if this formulation is not fully consistent with the smoother

theory, we will now investigate the possibility and the efficiency of retrospective analyses in a context of fixed-basis formulation of the filter. These results are exposed in the next section.

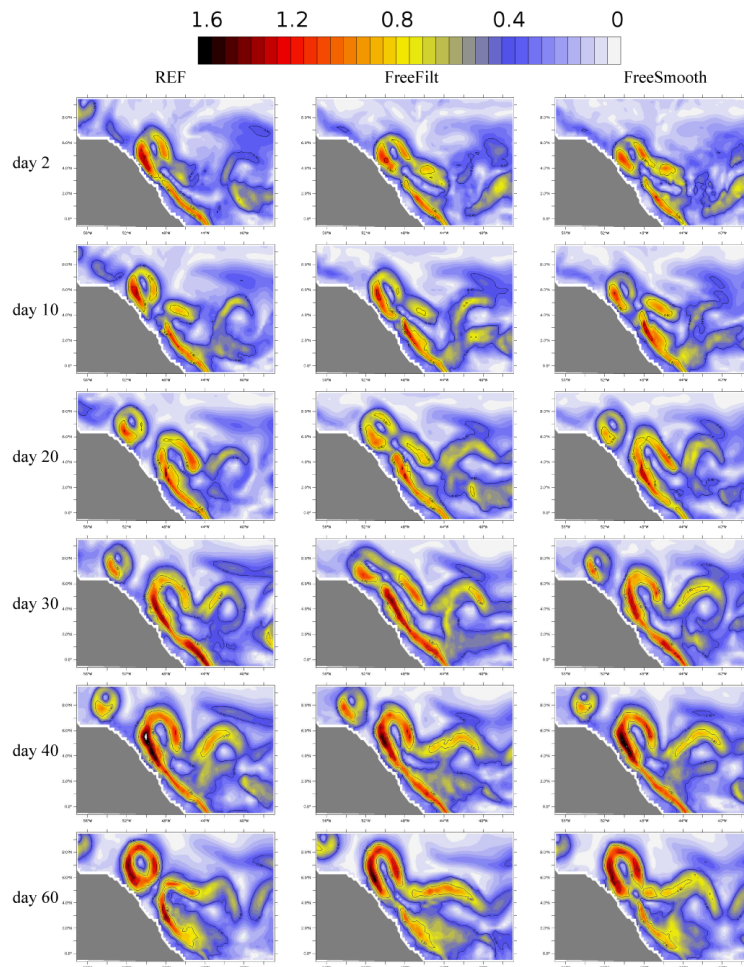
## 6 Smoother based on a static filter

It is well known that the main obstacle to accurate Kalman filtering in oceanography is model dimension. At the French ocean forecast service Mercator Ocean (<http://www.mercator-ocean.fr/fre/>), the model is presently run at a  $1/12^\circ$  resolution and 50 vertical levels, featuring more than a billion state variables and a prohibitive cost for ensemble methods. Consequently, data assimilation is performed with a rather simple optimal interpolation (OI) scheme, as at several other centers. In this section, we present and experiment a formulation of the fixed-lag smoother that complies with these operational constraints. The filter described and implemented previously is modified to work as an OI scheme. The same parameterisations are used, except that Eq. (25) is now replaced by the following equation:

$$\mathbf{S}_{k|k-1}^f = \mathbf{S}_0^a, \quad (26)$$

$\mathbf{S}_0^a$  being the initial basis of error modes.

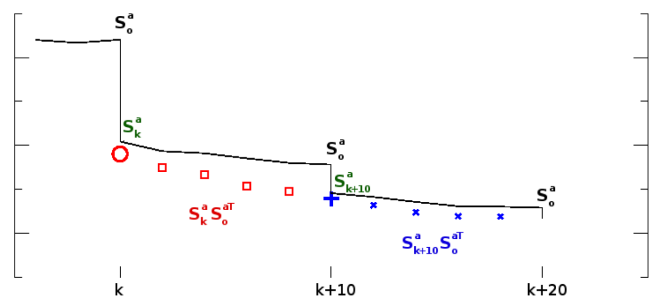
A hypothesis inherent to the implementation of an optimal interpolation scheme is that the analyses times are distant enough to make the residual errors decorrelate from one step to the next. To implement the OI scheme with our configuration, we define a new observation network with a 10-day frequency. Observations themselves are the same (i.e. they are extracted from REF every 2 days), but are gathered together at the nearest analysis days, and assimilated all together on



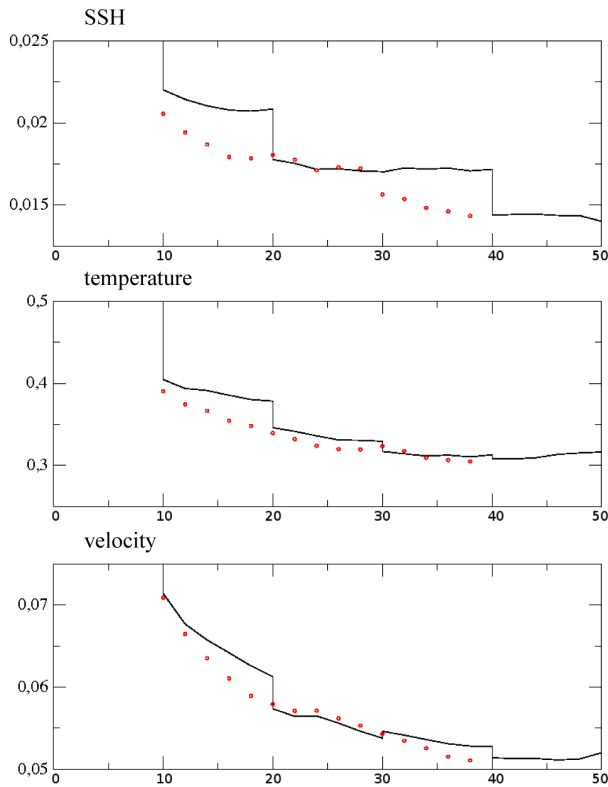
**Fig. 8.** Absolute error maps for the velocity (in  $\text{m s}^{-1}$ ), for REF (left), FreeFilt (middle) and FreeSmooth (right). Day 0 (top) is the day of the restart. The snapshots are given every 10 days.

those days. For instance, observations between day 5 and 15 are gathered and assimilated at day 10. The next assimilation step occurs at day 20, with all observations available between day 15 and 25. Day 10 is now the date of the first assimilation step.

Figure 9 presents a diagram of the smoother implementation. At the filter analysis time  $k$ , the analysis covariance matrix  $\mathbf{S}_k^a$  is computed from the background  $\mathbf{S}_0^a$ , simultaneously to the analysis state. This is not usually done with an OI scheme. The retrospective analysis at time  $k$  using the observation at a time  $k + 10$  is performed (corresponding to a 10-day interval) using the cross-covariances matrix  $\mathbf{S}_k^a \mathbf{S}_0^{aT}$ . Thus, the smoother uses information on residual errors at time  $k$ , combined with  $\mathbf{S}_0^a$  instead of  $\mathbf{S}_{k+10|k}^f$ . If  $\mathbf{S}_0^a$  is fixed, the matrix  $\mathbf{S}_k^a$  depends on the time of the analysis (this matrix is computed every 10 days after an OI analysis step). This form of smoother can be then considered as half-fixed. The large red circle and the blue plus sign in Fig. 9 represent the smoother analysis with a 10-day retrospective analysis.



**Fig. 9.** Schematic representation of the half-fixed basis smoother with a 10-day frequency observation network. Analysis covariance matrices (in green) are computed after each optimal interpolation analysis (at time  $k$  and  $k + 10$ ). The cross-covariance matrix used for the smoother analysis between date  $k$  and  $k + 8$  (respectively between  $k + 10$  and  $k + 20$ ) is indicated in red (respectively blue). Smoothed states are indicated with red or blue symbols (see Sect. 6 for more explanations).

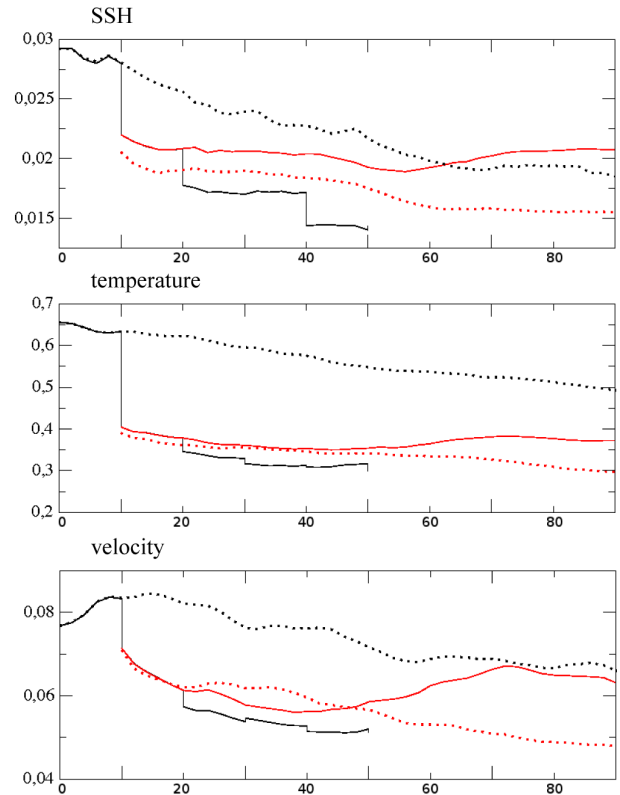


**Fig. 10.** RMS errors for the SSH (m), the temperature ( $^{\circ}\text{C}$ ) and the zonal velocity ( $\text{m s}^{-1}$ ), for the optimal interpolation (full black line) and the smoother (red circles).

Now, it is also possible to perform smoother analyses at intermediate times, every 2 days for instance. For each of such analyses, the background state is provided by the model simulation during the OI process. The background error modes must be defined appropriately, based on already existing bases. In the following experiments, the modes at times  $k+2, \dots, k+8$  are simply prescribed with  $\mathbf{S}_k^a$ , but other sound solutions are possible, such as a linear combination of  $\mathbf{S}_k^a$  and  $\mathbf{S}_0^a$ . This was not tested here. These intermediate smoother analyses are represented with red squares and blue crosses in Fig. 9.

### 6.1 Smoother's impact on the optimal interpolation solution

Results on RMS error levels for the optimal interpolation and the smoother based on OI are shown on Fig. 10. The smoother impact (red circles) is visible for each variable, assimilated or not. First, the evolution of error levels for the smoothed solution is continuous. Discontinuities inherent to sequential assimilation solutions (visible here every 10 days, during OI analysis) are smoothed out. Also, even with a very simple configuration of cross-covariance matrices, the smoothed solution exhibits lower error levels than the filtered one, except for some points (between day 20 and 30 for



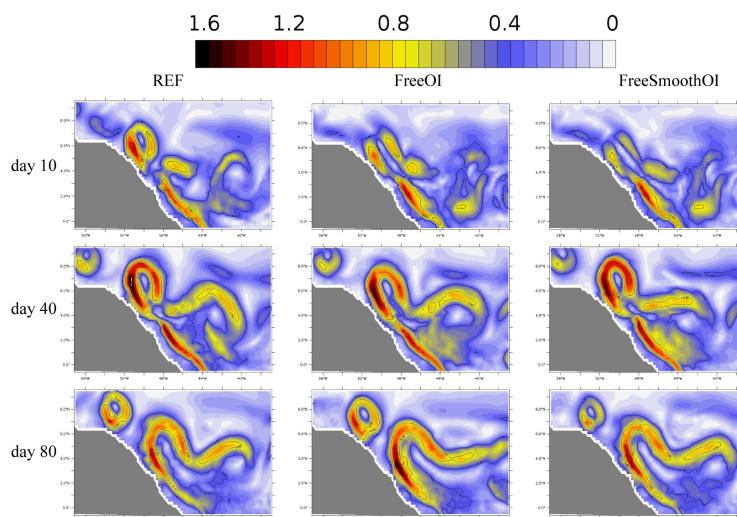
**Fig. 11.** RMS errors for the SSH (m), the temperature ( $^{\circ}\text{C}$ ) and the zonal velocity ( $\text{m s}^{-1}$ ), for the optimal interpolation (full black line) and the free runs restarted with the corrected state at day 10 with OI (full red line) or the smoothed state based on OI at day 10 (dotted red line). The RMS errors for FALSE are also indicated by the dotted black lines.

the velocity). It is finally worth noticing that this smoothing scheme is almost cost-free and could be applied with current assimilation systems in operational oceanographic centers.

Note that a 2-day configuration of a smoother can also be used with a 10-day frequency evolutive filter. The smoother needs the propagation of error statistics between the beginning of an interval and the day of the retrospective assimilation. This scheme is then similar to the asynchronous Kalman filter described by Hunt et al. (2004).

### 6.2 Dynamics with a half-fixed basis smoother

As in the previous section, we here examine the dynamical balance of the smoother solution. Free runs are initialized at day 10 with the analysed states from OI (FreeOI) or the smoother based on OI (FreeSmoothOI). The evolution of error levels are shown in Fig. 11. Even if initial error levels for OI and the smoother are close (especially for temperature and velocity), the behaviours of both free runs are significantly different. The error level of FreeOI tends to increase (after day 50) for each variable. On the contrary, for each variable, the error level of FreeSmoothOI tends to decrease



**Fig. 12.** Absolute error map for the velocity (in  $\text{m s}^{-1}$ ), for REF (left), the run restarted with a corrected state for OI (middle) and a smoothed state based on OI (right). Day 0 (top) is the day of the restart. The snapshots are given every 40 days.

continuously. This is blatant with velocity. Even after day 50, the error level keeps decreasing, suggesting that the dynamics stay consistent and close to REF. Figure 12 illustrates this point in the North Brazil Current region. During the first 40 days, the Brazil rings are quite similar between FreeOI and FreeSmoothOI. However, at day 80, the rings and the branch of the North Brazil Current's retroflexion are clearly better simulated in FreeSmoothOI.

This example shows once again that the improvement of the solution due to the smoother is not only statistical, but has a real dynamical meaning. The smoother can be used with a non-optimal system (fixed-basis) and still exhibits significant improvements, both on the average error level and on the dynamics of the solution. Still, one should keep in mind that the parametrisation of the OI here is not optimal, since it is based on the same parametrisation of an evolutive filter. Moreover, as already discussed, many possibilities can be considered using a smoother scheme based on OI. We chose a simple configuration here, but more sophisticated ways to define the cross-covariance error matrix could improve the results. These results also suggest that a smoother scheme could be introduced in current operational center systems, opening new perspectives for upcoming reanalyses of the ocean circulation.

## 7 Conclusions

The reduced-rank sequential smoother (Cosme et al., 2010) based on the SEEK filter has been further developed and implemented in a real ocean circulation model. The main purpose was to study the obstacles to implementing such an assimilation scheme in a real reanalysis system, and to study effects of the smoother on the estimation of the corrected

states. The smoother was first implemented as described in its seminal formulation, i.e. with an evolutive filter. In a second step, a simplified scheme was introduced based on optimal interpolation, as many operational centers use, and its benefits were exposed. A simple twin experiment was set up for this study, based on a perturbation of the initial condition. This configuration limited the benefits of the analysis throughout time, but a focus on the first assimilation steps enabled us to draw relevant conclusions about the smoother behaviour and benefits.

Technically, the implementation of the smoother algorithm is rather straightforward (when a filter algorithm is already implemented). It requires only a few extra calculations compared to the filter, and can be performed simultaneously to or after a filter pass. The smoother can then be considered as an additional layer to the filter algorithm.

A major obstacle to implementing a smoother lies in the parametrisation of an evolutive filter, theoretically needed to define accurately the cross-covariance error matrices. We used a criterion based on mean error statistics to specify the parameters associated with the filter. We found that for the mean error statistics of the filter to be consistent with the true error, we had to prescribe large observation errors. Moreover, the filter efficiency was mostly significant during the first 10 days of the experiment. To preserve optimality over a longer period, the filter should be used with adaptive parametrisations (Brankart et al., 2010).

To circumvent this issue, we considered implementing a smoother based on a simpler system of optimal interpolation. To do so, we defined a new algorithm for a half-fixed basis smoother. The advantage of such formulation is that it could be implemented almost as is with current assimilation systems of operational centers. This smoother was tested with a 10-day frequency observation network, but we also pointed



out the ability of the smoother to provide a solution with a higher frequency (2 days in this case).

The smoother efficiency was studied in both cases: evolutive and fixed error modes. Overall, the smoother was able to improve the results compared to 3-D assimilation schemes, filter or optimal interpolation. In this twin experiment, with absence of perturbations to the forcing and to the synthetic observations, error levels were reduced with the retrospective assimilation of observations. The smoother impact was mostly significant during the first 10 days of the experiment, just like the filter, underlying the fact that the smoother efficiency is strongly related to the parametrisation of the filter (and the optimal interpolation).

We also could verify that the smoothed solution was dynamically consistent and in better dynamical adequacy with the reference than the filtered solution. Thus, the corrections introduced with the smoother are not only statistical but also have a dynamical consistency. This point was surprisingly true in the case of the half-fixed basis smoother, which is encouraging in the context of using the smoother to improve the quality of reanalyses.

We suggest that the smoother could be performed with actual reanalysis systems. Indeed, the algorithm could be implemented easily and with a negligible cost, and the smoother has shown encouraging results with an optimal interpolation system. In other terms, there is no reason to not use it.

*Acknowledgements.* The authors thank the 2 anonymous reviewers and the editor who contributed to improving the initial manuscript. This work was jointly supported by the French Ministère de l'Enseignement Supérieur et de la Recherche, the Groupe Mission MERCATOR/CORIOLIS, the European Commission under Grant Agreement FP7-SPACE-2007-1-CT-218812-MYOCEAN, and CNES. The calculations were performed using HPC resources from GENCI-IDRIS (Grant 2009-011279).

Edited by: A. Schiller

## References

- Allen, M. R., Lawrence, S. P., Murray, M. J., Mutlow, C. T., Stockdale, T. N., Llewellynjones, D. T., and Anderson, D. L. T.: Control of tropical instability waves in the pacific, *Geophys. Res. Lett.*, 22, 2581–2584, 1995.
- Anderson, B. D. O. and Moore, J. B.: *Optimal Filtering*, Prentice-Hall, 357 pp., 1979.
- Brankart, J.-M., Testut, C.-E., Brasseur, P., and Verron, J.: Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: application to a 1993–96 hindcast of the North Atlantic Ocean circulation, *J. Geophys. Res.*, 108, 1–20, 2003.
- Brankart, J.-M., Ubelmann, C., Testut, C.-E., Cosme, E., Brasseur, P., and Verron, J.: Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry, *Mon. Weather Rev.*, 137, 1908–1927, 2009.
- Brankart, J.-M., Cosme, E., Testut, C.-E., Brasseur, P., and Verron, J.: Efficient adaptative error parametrizations for square root or ensemble Kalman filters: application to the control of ocean mesoscale signals, *Mon. Weather Rev.*, 138, 932–950, 2010.
- Brankart, J.-M., Cosme, E., Testut, C.-E., Brasseur, P., and Verron, J.: Efficient local error parameterizations for square root or ensemble Kalman filters: application to a basin-scale ocean turbulent flow, *Mon. Weather Rev.*, 139, 474–493, 2011.
- Brasseur, P. and Verron, J.: The SEEK filter method for data assimilation in oceanography: a synthesis, *Ocean Dynam.*, 56, 650–661, doi:10.1007/s10236-006-0080-3, 2006.
- Brasseur, P., Bahurel, P., Bertino, L., Birol F., Brankart, J.-M., Ferry, N., Losa, S., Remy, E., Schröter, J., Skachko, S., Testut, C.-E., Tranchant, B., Van Leeuwen, P. J., and Verron, J.: Data Assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments, *Q. J. Roy. Meteor. Soc.*, 131, 3561–3582, 2005.
- Brodeau, L., Barnier, B., Penduff, T., Treguier, A.-M., and Gulev, S.: An ERA-40 based atmospheric forcing for global ocean circulation models, *Ocean Model.*, 31, 88–104, 2010.
- Castruccio, F., Verron, J., Gourdeau, L., Brankart, J.-M., and Brasseur, P.: On the role of the GRACE mission in the joint assimilation of altimetry and TAO data in a tropical Pacific ocean model, *Geophys. Res. Lett.*, 33, L14616, doi:10.1029/2006GL025823, 2006.
- Cohn, S. E., Sivakumaran, N. S., and Todling, R.: A fixed-lag Kalman smoother for retrospective data assimilation, *Mon. Weather Rev.*, 122, 2838–2867, 1994.
- Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, *Ocean Model.*, 33, 87–100, 2010.
- Cosme, E., Verron, J., Brasseur, P., Blum, J., and Auroux, D.: Smoothing problems in a Bayesian framework and their linear Gaussian solutions, *Mon. Weather Rev.*, 140, 683–695, 2011.
- Dussin, R., Treguier, A.-M., Molines, J., Barnier, B., Penduff, T., Brodeau, L., and Madec, G.: Definition of the interannual experiment orca025-b83, Tech. rep., 2009.
- Evensen, G.: Sequential data assimilation with a non linear quasi-geostrophic model using monte carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, 1994.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003.
- Evensen, G.: *Data assimilation, The Ensemble Kalman Filter*, Springer, 2007.
- Evensen, G. and van Leeuwen, P. J.: An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, 128, 1852–1867, 2000.
- Ferry, N., Parent, L., Garric, G., Barnier, B., Jourdin, N. C., and Mercator Ocean team: Mercator Global Eddy Permitting Ocean Reanalysis GLORYS1V1: Description and Results, *Mercator Ocean Quarterly Newsletter*, 36, 15–27, 2010.
- Fukumori, I. and Malanotte-Rizzoli, P.: An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf stream model, *J. Geophys. Res.*, 100, 6777–6793, 1995.
- Garzoli, S. and Katz, E.: The forced annual reversal of the north atlantic equatorial countercurrent, *J. Phys. Oceanogr.*, 13, 2082–2090, 1993.



- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an Ensemble Kalman Filter technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L., Mitchell, H. L., and Deng, X.: Model error representation in an operational Ensemble Kalman Filter, *Mon. Weather Rev.*, 137, 2126–2143, 2009.
- Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, *Tellus A*, 56, 273–277, 2004.
- Kalman, R. E.: A new approach to linear filter and prediction problems, *J. Basic. Eng.*, 82, 35–45, 1960.
- Khare, S. P., Anderson, J. L., Hoar, T. J., and Nychka, D.: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems, *Tellus A*, 60, 97–112, 2008.
- Legeckis, R.: Long waves in the eastern equatorial Pacific ocean: a view from a geostationary satellite, *Science*, 197, 1179–1181, 1977.
- Lermusiaux, P. F. J., Robinson, A. R., Haley, P. J. H., and Leslie, W. G.: Advanced interdisciplinary data assimilation: filtering and smoothing via error subspace statistical estimation, in: *Proceedings of the OCEANS 2002 MTS/IEEE Conference*, IEEE, Holland Publications, 795–802, 2002.
- Li, H., Kalnay, E., and Miyoshi, T.: Simultaneous estimation of covariance inflation and observation errors within an Ensemble Kalman Filter, *Q. J. Roy. Meteor. Soc.*, 135, 523–533, 2009.
- Liu, Z. and Rabier, F.: The interaction between model resolution and observation resolution and density in data assimilation, *Q. J. Roy. Meteor. Soc.*, 128, 1367–1386, 2002.
- Madec, G.: Nemo reference manual, ocean dynamics component: nemo-opa. preliminary version, Tech. Rep., Institut Pierre-Simon Laplace (IPSL), France, 27, ISSN No. 1288–1619, 91 pp., available at: <http://www.locean-ipsl.upmc.fr/NEMO/general/manual/index.html>, 2008.
- Martin, M. J., Hines, A., and Bell, M. J.: Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact, *Q. J. R. Meteor. Soc.*, 133, 981–995, 2007.
- Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A.: The Bluelink ocean data assimilation system (BODAS), *Ocean Model.*, 21, 46–70, 2008.
- Parrish, D. F. and Cohn, S. E.: A Kalman filter for a two-dimensional shallow-water model: formulation and preliminary experiments, Tech. Rep., Office Note 304, National Meteorological Center, Washington DC, 1985.
- Pham, D. T., Verron, J., and Roubaud M. C.: A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar. Syst.*, 16, 323–340, 1998.
- Simon, D.: *Optimal State Estimation*, Wiley & Sons, 530 pp., 2006.
- Rabier, F.: Importance of data: A meteorological perspective, in: *Ocean Weather Forecasting: An Integrated View of Oceanography*, edited by: Chassignet, E. P. and Verron, J., Springer, 343–360, 2006.
- Richardson, P. and Walsh, D.: Mapping climatological seasonal variations of surface currents in the tropical Atlantic using ship drifts, *J. Geophys. Res.*, 91, 10537–10550, 1986.
- Rozier, D., Cosme, E., Birol, F., Brasseur, P., Brankart, J.-M., and Verron, J.: A reduced-order Kalman filter for data assimilation in physical oceanography, *SIAM Rev.*, 49, 449–465, 2007.
- Testut, C.-E., Brasseur, P., Brankart, J.-M., and Verron, J.: Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy permitting primitive equation model of the North Atlantic Ocean, *J. Mar. Syst.*, 40–41, 291–316, 2003.
- Todling, R. and Cohn, S. E.: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter, *Mon. Weather Rev.*, 122, 2530–2557, 1994.
- Ubelmann, C., Verron, J., Brankart, J.-M., Brasseur, P., and Cosme, E.: Impact of data from upcoming altimetric missions on the control of the three dimensional circulation in the tropical Atlantic ocean, *J. Operat. Oceanogr.*, 2, 3–14, 2009.
- Verron, J., Gourdeau, L., Pham, D. T., Murtugudde, R., and Busalacchi, A. J.: An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the Tropical Pacific: method and validation, *J. Geophys. Res.*, 104, 5441–5458, 1999.