Manuscript prepared for Nonlin. Processes Geophys.
with version 4.2 of the LaTeX class copernicus.cls.
Date: 6 March 2015

Data assimilation Experiments using the Diffusive Back and Forth Nudging for the NEMO ocean model

Ruggiero, G. A.[1], Ourmières, Y.[2], Cosme, E.[3], Blum, J.[1], Auroux, D.[1], and Verron, J.[4]

[1]Université de Nice Sophia-Antipolis/LJAD, Nice, France
[2] Université du Sud Toulon-Var, Aix-Marseille Université, CNRS/INSU, IRD, Mediterranean Institute of Oceanography (MIO), France
[3]Université Joseph Fourier/LGGE, Grenoble, France
[4]CNRS/LGGE, Grenoble, France

Correspondence to: Giovanni A. Ruggiero
giovanni.ruggiero@unice.fr

Abstract. The Diffusive Back and Forth Nudging (DBFN) is an easy-to-implement iterative data assimilation method based on the well-known Nudging method. It consists in a sequence of forward and backward model integrations, within a given time window, both of them using a feedback term to the observations. Therefore in the DBFN, the Nudging asymptotic behavior is translated into an infinite number of iterations within a bounded time domain. In this method, the backward integration is carried out thanks to what is called backward model, which is basically the forward model with reversed time step sign. To maintain numeral stability the diffusion terms also have their sign reversed, giving a diffusive character to the algorithm. In this article the DBFN performance to control a primitive equation ocean model is investigated. In this kind of model non-resolved scales are modeled by diffusion operators which dissipate energy that cascade from large to small scales. Thus, in this article the DBFN approximations and their consequences on the data assimilation system set-up are analyzed. Our main result is that the DBFN may provide results which are comparable to those produced by a 4Dvar implementation with a much simpler implementation and a shorter CPU time for convergence. The conducted sensitivity tests show that the 4Dvar profits of long assimilation windows to propagate surface information downwards, and that for the DBFN, it is worth using short assimilation windows to reduce the impact of diffusion-induced errors. Moreover, the DBFN is less sensitive to the first guess than the 4Dvar.

Keywords. Data Assimilation, Nudging, Back and Forth Nudging, NEMO

1 Introduction

In data assimilation, an interesting tool is the Kalman-Bucy filter (Kalman and Bucy, 1961), where a non-linear differential equation of the Riccati type was derived for the covariance matrix of the optimal filtering error, the solution of which completely specifies the optimal filter for linear-quadratic problems. A few years later, Luenberger (1966, 1971) defined an observer for reconstructing the state of an observable deterministic linear system from exact measurements of the output. This Luenberger observer has been called "asymptotic estimator", since under linearity and observability hypothesis the estimator error converges to zero for time tending to infinity (Gelb et al., 1974; Bonnans and Rouchon, 2005). Its advantage compared to Kalman filtering is that it does not require any information on the various covariance matrices, but as it was pointed out in Luenberger (1966), the Kalman-Bucy filter appears as a particular Luenberger observer which is the optimal least-mean-square estimate of the state in the case of noisy measurements. The stochastic observer unifies the concepts of deterministic Luenberger observer theory and stochastic Kalman filtering theory as it is explained in Gelb's book (Gelb et al., 1974) for instance. Both are useful in practice. It should be mentioned that the concept of Luenberger observer has been extended to nonlinear systems for example in Zeitz (1987).

This Luenberger observer has been rediscovered in the geophysical literature for atmospheric models under the term of nudging (Anthes, 1974; Hoke and Anthes, 1976; Stauffer and Seaman, 1990), which consists in adding a forcing term in the right hand side of a given model in order to gently push (nudge) the solution toward a prescribed value. It is quite interesting to note that there is no mention of the link between nudging and Luenberger observer in the geophysical literature until the work of Auroux and Blum (2008). More recently, a comprehensive study on the nudging method and its variants was produced by Blum et al. (2008) and Lakshmivarahan and Lewis (2012).

The first appearance of a successful application of nudging to ocean Data Assimilation (DA) was in 1992 in a work that assimilated sea surface height derived from satellite measurements into a quasi-geostrophic layered model (Verron, 1992). Since then, the method has been successfully applied to several oceanographic numerical problems such as the estimation of boundary conditions (Marchesiello et al., 2001; Chen et al., 2013), downscaling (Li et al., 2012), and other DA problems (Verron, 1992; Haines et al., 1993; Blayo et al., 1994; Lewis et al., 1998; Killworth et al., 2001; Thompson et al., 2006). Concerning applications to DA problems, the weights given to the model and the observations are generally not based on any optimality condition, but are rather scalars or Gaussian-like functions constructed based on physical assumptions or empirical considerations. The appeals of this method are the simplicity of implementation in complex numerical models, the low computational power required and the time smoothness of the solution.

The increasing availability of computing power has allowed to use more advanced data assimilation methods. In general, these methods use information on the model statistics and observations errors to weight the model-observations combination. Two of these methods that are widely used by prediction centers are the ensemble Kalman filter- EnKF (Evensen, 1994) and its variations (Pham, 2001; Hunt et al., 2007), and the four dimensional variational method 4Dvar (Le Dimet and Talagrand, 1986; Courtier et al., 1994). For the first, the numerical costs are due to the propagation of the ensemble, usually formed by tenths of members, to calculate the forecast. For the second, the costs are due to the need of minimizing a cost function in a very large state space (10^8 variables). This requires several iterations of the minimization algorithm, which involves several integrations of the direct and adjoint models.

However, even with the growing interest in these complex techniques built on solid theoretical arguments, nudging has not been left aside. Recent works have used nudging along with more advanced methods such as Optimal interpolation (Clifford et al., 1997; Wang et al., 2013), EnKF (Ballabrera-Poy et al., 2009; Bergemann and Reich, 2010; Lei et al., 2012; Luo and Hoteit, 2012), 4Dvar (Zou et al., 1992; Stauffer and Bao, 1993; Vidard et al., 2003; Abarbanel et al., 2010) or particle filters (Luo and Hoteit, 2013; Lingala et al., 2013) to extract the best of each method. In the particular case of the hybridization with the EnKF proposed by Lei et al. (2012), the resulting algorithm takes the advantage of the dynamical propagation of the covariance matrix from the EnKF and uses nudging to mitigate problems related to the intermittence of the sequential approach, which among other things entails the possible discarding of some observations.

In 2005, Auroux and Blum (2005) revisited the nudging method and proposed a new observer called Back and Forth Nudging (BFN), because Luenberger observer is an asymptotic observer, and as data assimilation is performed for finite time, the convergence of the real state is not yet achieved at limited horizon. The BFN consists in a sequence of forward and backward model integrations, both of them using a feedback term to the observations, as in the direct nudging. The BFN integrates the direct model backwards in time avoiding the construction of the adjoint and/or tangent linear models needed by 4DVar. Therefore, it uses only the fully non-linear model to propagate information forward and backward in time. The nudging gain, which has an opposite sign with respect to the forward case, has a double role: push the model toward observations and stabilize the backward integration, which is especially important when the model is not reversible. Back and forth algorithms have already been used in the past for initialization and four-dimensional data assimilation (Morel et al., 1971; Talagrand, 1981), but without nudging. In these papers, the authors are just replacing at each observation time the values predicted by the model for the observed parameters by the observed values; this method requires the considered system to be reversible, which is not the case if there exists irreversible dissipation in the model.

The BFN convergence was proved by Auroux and Blum (2005) for linear systems of ordinary differential equations and full observations, by Ramdani et al. (2010) for reversible linear partial differential equations (Wave and Schrödinger equations), by Donovan et al. (2010) and Leghtas et al. (2011) for the reconstruction of quantum states and was studied by Auroux and Nodet (2012) for non-linear transport equations. The BFN performance in numerical applications using a variety of models, including non-reversible models such as a Shallow Water (SW) model (Auroux, 2009) and a Multi-Layer Quasi-Geostrophic (LQG) model (Auroux and Blum, 2008), are very encouraging. Moreover, by using a simple scalar gain, it produced results comparable to those obtained with 4DVar but with lower computational requirements (Auroux, 2009; Auroux et al., 2012).

In this article we present for the first time a BFN application to control a primitive equation ocean model. The numerical model used is NEMO (Madec, 2008), currently used by the French operational center, Mercator Océan (http://www.mercator-ocean.fr/fre), to produce and deliver ocean forecasts. The well-known idealized double gyre configuration at eddy-permitting resolution is used. This configuration has the advantage of being simple from the geometry and forcings point of view at the same time it reproduces most of features found in a middle latitude ocean basin.

The BFN application to control a primitive equation ocean model represents a new challenge due to the increased model complexity. Among the differences between NEMO and the simplified oceanic models used by Auroux and Blum (2008) and Auroux (2009) stand out the more complex relationship between the variables in the former since no filtering technique is used in the derivation of the physical model (except the Boussinesq approximation which is also considered by the SW and LQG models), and the inclusion of an equation for the conservation of the thermodynamical properties. The latter requires the use of a nonlinear state equation to couple dynamical and thermodynamical variables.

Furthermore, the vertical ocean structure represented by NEMO is more complex than the vertical ocean structure represented by the SW and LQG used by Auroux and Blum (2008) and Auroux (2009). This is because the SW model has no vertical levels and the LQG was implemented with only 3 layers, while in this article NEMO is configured with 11 vertical layers. In addition, NEMO considers vertical diffusion processes, mostly ignored by the LQG model. Vertical diffusion plays an important role in maintaining the ocean stratification and meridional overturning circulation, which is directly related to the transport of heat in the ocean. Moreover from the practical point of view, the diffusion/viscosity required to keep the NEMO simulations stable is by far greater than for the SW or LQG at the same resolution.

These issues call into question the validity of the approximations made by the BFN under realistic conditions. Thus, our primary objective is to study the possibility of applying the BFN in realistic models and evaluate its performance compared to the 4Dvar. This appears as being the next logical step before using the BFN to assimilate real data.

This article is organized as follows. In Sect 2 the BFN and the 4Dvar are described. Section 3 describes the model physics and the model set-up. Section 4 discusses some practical aspects of the backwards integration. Section 5 presents the BFN and the 4Dvar set-up and the designed data assimilation experiments. Finally, the data assimilation results are presented in the Sect 6, on which we discuss the impact of the length of the data assimilation window on the method performances as well as the sensitivity of each method to the observation network and the initial condition.

2 Data Assimilation Methods

In this section the Back and Forth Nudging (BFN) is introduced and the 4Dvar used to assess the BFN performance is briefly described.

2.1 The Back and Forth Nudging

The conventional nudging algorithm consists in adding a forcing term (feedback term) to the model equations, proportional to the difference between the data and the model at a given time. More generally, given a model described by a set of ordinary equations (or discretized partial differential equations), nudging consists in adding to them the forcing term $\boldsymbol{K}(\boldsymbol{x}_{obs} - \mathcal{H}(\boldsymbol{x}))$:

$$\frac{d\boldsymbol{x}}{dt} = \mathcal{F}(\boldsymbol{x}) + \boldsymbol{K}(\boldsymbol{x}_{obs} - \mathcal{H}(\boldsymbol{x})) \qquad (1)$$

where $\boldsymbol{x}$ represents the state vector, $\mathcal{F}$ is the model operator, $\mathcal{H}$ is the observation operator allowing one to compare the observations $\boldsymbol{x}_{obs}(t)$ to the corresponding system state $\mathcal{H}(\boldsymbol{x})$, and $\boldsymbol{K}$ is the nudging gain matrix. In this algorithm the model appears as a weak constraint. The feedback term changes the dynamical equations and is a penalty term that forces the state variables to get closer to the observations.

In the linear case, i.e. when $\mathcal{F}$ and $\mathcal{H}$ may be written as matrices $\boldsymbol{F}$ and $\boldsymbol{H}$, and in the absence of noise in the system, nudging is nothing else than the Luenberger observer (Luenberger, 1966). In this case, and assuming that the observability of the pair $(\boldsymbol{F},\boldsymbol{H})$ holds, there is a class of possible matrices $\boldsymbol{K}$ that, thanks to the pole shifting theorem, guarantees the estimator convergence when $t \rightarrow \infty$ (Gelb et al., 1974; Bonnans and Rouchon, 2005). This should be one possible explanation why nudging usually works quite well and the converged state is not strongly affected by the choice of $\boldsymbol{K}$. However, when constructing $\boldsymbol{K}$ (which units is s^{-1}), the aim is to obtain an estimator response faster than the time scale of the studied processes.

The BFN is an iterative algorithm which sequentially solves the forward model equations with a feedback term to the observations (Eq. 1) and the backward model equations with an opposite sign for the feedback term. The initial condition of the backward integration is the final state obtained after integration of the forward nudging equation. At the end of each iteration a new estimation of the system's initial state is obtained. The iterations are carried out until convergence is reached.

The difference of the BFN with respect to the conventional nudging is the model integration backward in time. This allows to recover initial conditions as well as to use more than once the same observations set. Consequently, the BFN may be seen as a sub-optimal iterative smoother.

Under the hypothesis of a linear model a variational interpretation is possible. In this case, if we choose $\boldsymbol{K} = k\boldsymbol{H}^T\boldsymbol{R}^{-1}$, where $\boldsymbol{R}$ is the observation error covariance matrix, and k is a scalar, the solution of the forward nudging is a compromise between the minimization of the system's energy and the minimization of the distance between the data and the model (Auroux and Blum, 2008).

However, the backward integration is problematic when the model is diffusive or simply not reversible. In the case of ocean models, there are two main aspects requiring the inclusion of diffusion: i) the control of numerical noise, and ii) the modeling of sub grid-scale processes, i.e. to parameterize the energy transfer from explicitly resolved to non-resolved scales. Indeed, diffusion naturally represents a source of uncertainty in ocean forecasts, even for the purely forward model, and has been investigated from the point of view of the optimal control theory in Leredde et al. (1999).

To address the problem of the backward model instability in this article the Diffusive Back and Forth Nudging-DBFN (Auroux et al., 2011) is used. In this algorithm the sign of the diffusion term remains physically consistent and only the reversible part of the model equations are really solved backward. Practical consequences of this assumption are analysed in Sect 4. A similar solution was proposed by Pu et al. (1997) and Kalnay et al. (2000) to stabilize their Quasi-Inverse Linear model.

To describe the DBFN algorithm, let us assume that the time continuous model satisfies dynamical equations of the form:

$$\frac{\partial \boldsymbol{x}}{\partial t} = \mathcal{F}(\boldsymbol{x}) + \nu \Delta \boldsymbol{x}, \qquad \text{for} \qquad 0 < t < T, \tag{2}$$

with an initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$, where $\mathcal{F}$ denotes the nonlinear model operator without diffusive terms, ν is a diffusion coefficient and Δ represents a diffusion operator. If nudging is applied to the forward system (2) it gives:

$$\begin{aligned} \frac{\partial \boldsymbol{x}_k}{\partial t} &= \mathcal{F}(\boldsymbol{x}_k) + \nu \Delta \boldsymbol{x}_k + \boldsymbol{K}(\boldsymbol{x}_{obs} - \mathcal{H}(\boldsymbol{x}_k)) \\ \boldsymbol{x}_k(0) &= \tilde{\boldsymbol{x}}_{k-1}(0), \qquad 0 < t < T, \end{aligned} \tag{3}$$

where $k \in \mathbb{N}_{\geq 1}$ stands for iterations and $\tilde{\boldsymbol{x}}_0(0)$ is a given initial guess. Nudging applied to the backward system with the reversed diffusion sign gives:

$$\begin{aligned} \frac{\partial \tilde{\boldsymbol{x}}_k}{\partial t} &= \mathcal{F}(\tilde{\boldsymbol{x}}_k) - \nu \Delta \tilde{\boldsymbol{x}}_k - \boldsymbol{K'}(\boldsymbol{x}_{obs} - \mathcal{H}(\tilde{\boldsymbol{x}}_k)) \\ \tilde{\boldsymbol{x}}_k(T) &= \boldsymbol{x}_k(T), \qquad T > t > 0. \end{aligned} \tag{4}$$

The system composed by equations (3) and (4) is the basis of the DBFN algorithm. They are iterated until convergence.

Therefore, one important aspect of the DBFN algorithm is the convergence criterion. Ideally, at convergence the nudging term should be null or small comparable to the other equation terms. Otherwise, when the nudging is switched off, which is the case in the forecast phase, the system may return to a state close to the background state or to a state which is not consistent to the one at convergence. The convergence is calculated as:

$$\frac{\|\boldsymbol{x}_k(t=0) - \boldsymbol{x}_{k-1}(t=0)\|}{\|\boldsymbol{x}_{k-1}(t=0)\|} \leq \epsilon, \tag{5}$$

where $\| \bullet \|$ is the L_2 norm, and the choice for $\epsilon = 0.005$ is based on sensitivity tests (not presented in this article).

Data Assimilation is the ensemble of techniques combining the mathematical information provided by the equations of the model and the physical information given by the observations in order to retrieve the state of a flow. In order to show that the DBFN algorithm achieves this double objective, let us give a formal explanation of the way DBFN proceeds. If $\boldsymbol{K'} = \boldsymbol{K}$ and the forward and backward limit trajectory are equal, i.e $\tilde{\boldsymbol{x}}_\infty = \boldsymbol{x}_\infty$, then taking the sum between Eqs.(3) and (4) shows that $\boldsymbol{x}_\infty$ satisfies the model equations without diffusion:

$$\frac{\partial \boldsymbol{x}_\infty}{\partial t} = \mathcal{F}(\boldsymbol{x}_\infty) \tag{6}$$

while taking the difference between Eqs.(3) and (4) shows that $\boldsymbol{x}_\infty$ satisfies the Poisson equation:

$$\Delta \boldsymbol{x}_\infty = -\frac{\boldsymbol{K}}{\nu}(\boldsymbol{x}_{obs} - \mathcal{H}(\boldsymbol{x}_\infty)) \tag{7}$$

which represents a smoothing process on the observations for which the degree of smoothness is given by the ratio $\frac{\nu}{\boldsymbol{K}}$ (Auroux et al., 2011). Equation (7) corresponds, in the case where $\mathcal{H}$ is a matrix $\boldsymbol{H}$ and $\boldsymbol{K} = k\boldsymbol{H}^T\boldsymbol{R}^{-1}$, to the Euler equation of the minimization of the following cost-function

$$\mathcal{J}(\boldsymbol{x}) = k < \boldsymbol{R}^{-1}(\boldsymbol{x}_{obs} - \boldsymbol{H}\boldsymbol{x}), (\boldsymbol{x}_{obs} - \boldsymbol{H}\boldsymbol{x}) > + \nu \int_\Omega \|\nabla \boldsymbol{x}\|^2 \tag{8}$$

where the first term represents the quadratic difference to the observations and the second one is a first order Tikhonov regularisation term over the domain of resolution Ω. The vector $\boldsymbol{x}_\infty$, solution of (7), is the point where the minimum of this cost-function is reached. It is shown in Sect 6.1 that at convergence the forward and backward trajectories are very close, which justifies this qualitative justification of the algorithm.

The description of the used $\boldsymbol{K}$ matrix is given in the Sect 5.1.

2.2 Four Dimensional Variational Method - 4DVar

Variational methods minimize a cost function that measures the distance between the estimated state and the available observations. Let us assume that observations are available at every instant $(t_i)_{1 \le i \le N}$. Given a first guess $\boldsymbol{x}^b$ of the initial state, the 4DVar algorithm will find an optimal initial condition that minimizes the distance between the model trajectory and the observations in a given assimilation window. This optimal state is found by minimizing the following cost function:

$$\begin{aligned} J(\boldsymbol{x}_0) = & \frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{x}^b)^T \boldsymbol{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}^b) \\ & + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i[\mathcal{M}_{0,i}(\boldsymbol{x}_0)] - \boldsymbol{y}_i)^T \boldsymbol{R}_i^{-1}(\mathcal{H}_i[\mathcal{M}_{0,i}(\boldsymbol{x}_0)] - \boldsymbol{y}_i) \end{aligned} \tag{9}$$

where $\boldsymbol{B}$ is the background error covariance matrix and $\mathcal{M}_{0,i}$ represents the model integration from time t_0 to time t_i. $\boldsymbol{R}_i$,$\mathcal{H}_i$ and $\boldsymbol{y}_i$ are the observations error covariance matrix, the observation operator and the available observations at time t_i, respectively.

The optimal initial state is found by solving:

$$\nabla J(\boldsymbol{x}^a(t_0)) = 0 \tag{10}$$

The calculation of this gradient is done using the adjoint method proposed by Lions (1971) and brought to the meteorological context by Le Dimet and Talagrand (1986).

Since for ocean applications $\mathcal{M}$ and $\mathcal{H}$ are nonlinear, we used the incremental approach proposed by Courtier et al. (1994) which consists in solving a sequence of linearized quadratic problems, expecting this sequence would converge to the solution of the problem given by (9) and (10). In this case, the cost function will not be minimized with respect to the initial state but with respect to an increment $\delta \boldsymbol{x}_0$ defined by $\boldsymbol{x}_0 = \boldsymbol{x}^b + \delta \boldsymbol{x}_0$. The operators $\mathcal{H}$ or $\mathcal{M}$ are linearized in a neighborhood of $\boldsymbol{x}^b$ as:

$$\mathcal{M}_{0,i}(\boldsymbol{x}^b + \delta \boldsymbol{x}_0) \approx \mathcal{M}_{0,i}(\boldsymbol{x}^b) + \boldsymbol{M}_{0,i} \delta \boldsymbol{x}_0 \qquad \forall i \tag{11}$$

$$\mathcal{H}_i(\boldsymbol{x}^b + \delta \boldsymbol{x}_0) \approx \mathcal{H}_i(\boldsymbol{x}^b) + \boldsymbol{H}_i \delta \boldsymbol{x}_0 \qquad \forall i \tag{12}$$

and the new cost function is given by:

$$J(\delta \boldsymbol{x}_0) = \frac{1}{2} \delta \boldsymbol{x}_0^T \boldsymbol{B}^{-1} \delta \boldsymbol{x}_0 + \frac{1}{2} \sum_{i=0}^{N} (\boldsymbol{H}_i \boldsymbol{M}_{0,i} \delta \boldsymbol{x}_0 - \boldsymbol{d}_i)^T \boldsymbol{R}_i^{-1} (\boldsymbol{H}_i \boldsymbol{M}_{0,i} \delta \boldsymbol{x}_0 - \boldsymbol{d}_i) \tag{13}$$

where $\boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\boldsymbol{x}_b))$ is called the innovation vector. To take advantage of the nonlinearities it is a common practice to re-linearise $\mathcal{H}$ and $\mathcal{M}$ around a new model trajectory after some iterations of the minimizer. This new model trajectory is computed by integrating the nonlinear model forward in time using $\boldsymbol{x}_0^k = \boldsymbol{x}^b + \delta \boldsymbol{x}_0^k$ as initial condition, where k refers to the new run of the nonlinear model and $\delta \boldsymbol{x}_0^k$ is the increment previously obtained through the minimization of (13). This gives rise to what is called the inner loop and outer loop iterations. The algorithm implemented in NEMO, called NEMOVAR (Mogensen et al., 2009), uses this technics. It can be summarized as follows:

–**Initialisation** : $\boldsymbol{x}_0^0 = \boldsymbol{x}^b$

–**While** $k \leq k_{max}$ or $\|\delta \boldsymbol{x}_0^{a,k}\| > \epsilon$ (**Outer Loop**)

Do

•$\boldsymbol{d}_i^k = \boldsymbol{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\boldsymbol{x}_0^k))$

•Search the $\delta \boldsymbol{x}_0^{a,k}$ that minimises (**Inner Loop**):

$$\begin{aligned} J(\delta \boldsymbol{x}_0^k) = {} & \frac{1}{2} (\delta \boldsymbol{x}_0^k)^T \boldsymbol{B}^{-1} (\delta \boldsymbol{x}_0^k) \\ & + \frac{1}{2} \sum_{i=0}^{N} (\boldsymbol{H}_i \boldsymbol{M}_{0,i} \delta \boldsymbol{x}_0^k - \boldsymbol{d}_i^k)^T \boldsymbol{R}_i^{-1} (\boldsymbol{H}_i \boldsymbol{M}_{0,i} \delta \boldsymbol{x}_0^k - \boldsymbol{d}_i^k) \end{aligned} \tag{14}$$

•$\boldsymbol{x}_0^{k+1} = \boldsymbol{x}_0^k - \delta \boldsymbol{x}_0^{a,k}$

The description of the matrices $\boldsymbol{B}$ and $\boldsymbol{R}$ is given in the Sect 5.2.

3 Ocean Model and Experimental set-up

The ocean model used in this study is the ocean component of NEMO (Nucleus for European Modelling of the Ocean; Madec, 1996). This model is able to represent a wide range of ocean motions, from the basin scale up to the regional scale. Currently, it has been used in operational mode by the

French Mercator Océan group (http://www.mercator-ocean.fr) and the European Center for Medium Range Weather Forecast (ECMWF).

The model solves a system of five prognostic equations, namely the momentum balance for the horizontal velocities, an equation describing the evolution of the free surface, and the heat and salt conservation equations. A nonlinear equation of state couples the two tracers to the fluid fields. In this study, a linear free surface formulation is used along with the approach developed by Roullet and Madec (2000) to filter out the external gravity waves.

Equations are discretized using spherical coordinates in a Arakawa C grid. The model advances in time using a leap-frog scheme for all terms except for the vertical diffusive terms, which are treated implicitly. At every time step the model uses a Robert-Asselin (RA) temporal filter to damp the computational mode. The leap-frog scheme followed by the RA filter leads to a first order temporal scheme (Willians, 2009). Spatial discretization uses a centered second order formulation for both the advective and the diffusive terms.

The double gyre configuration, extensively used to study jet instabilities (Chassignet and Gent, 1991; Primeau, 1998; Chang et al., 2001), meso and submeso-scale dynamics (Levy et al., 2010) and data assimilation methods (Molcard et al., 2004; Krysta et al., 2011; Cosme et al., 2010), is used for the present study. The double gyre configuration simulates the ocean middle latitude dynamics and has the advantage of being simple, when compared to real applications, but still considering full dynamics and thermodynamics.

In our experiments we use a homogeneous horizontal grid with a 25km resolution and a vertical resolution ranging from 100m near the upper surface to 500m near the bottom. The bottom topography is flat and the lateral boundaries are closed and frictionless. The only forcing term considered is a constant wind stress of the form $\tau = \left(\tau_0 cos\left(\frac{2\pi(y-y_0)}{L}\right), 0\right)$, where y is the latitude geographic coordinate with $y_0 = 24^o$ and $y_0 \leq y \leq 44^o$, $L = 20^o$ and $\tau_0 = -0.1 N/m^2$. Horizontal diffusion/viscosity are modeled by a bilaplacian operator meanwhile a laplacian operator is used in the vertical. They all use constant coefficients in time and space: $\nu_h^{u,v} = -8 \times 10^{10} m^4/s$ and $\nu_v^{u,v} = 1.2 \times 10^{-4} m^2/s$ for the momentum equations and $\nu_h^{t,s} = -4 \times 10^{11} m^4/s$ and $\nu_v^{t,s} = 1.2 \times 10^{-5} m^2/s$ for temperature and salinity. The initial condition is similar to that used by Chassignet and Gent (1991) and consists of a homogeneous salinity field of 35psu and a temperature field created to provide a stratification which has a first baroclinic deformation radius of 44.7km. Velocity and sea surface height (SSH) fields are initially set to zero.

This double gyre configuration is currently used as the NEMO data assimilation demonstrator and as the experimentation and training platform for data assimilation activities (Bouttier et al., 2012). For the present work, the model was integrated for 70 years, in order to reach the statistical steady state. Afterwards, ten years of free model run were performed, that were used to calculate the regression models which are used to calculate the nudging matrix $\boldsymbol{K}$ (see Sect 5.1), and then two additional years were finally completed to be used as the truth, from which the observations were extracted.

4 The backward integration without Nudging: Practical aspects

The backward model uses exactly the same numerical scheme as the forward model. Since most of the model is solved using centered finite differences, the inverse version of the discretized model is similar to the discrete version of the inverse continuous model. The only distinction between the forward and the backward model is the change in the sign of the diffusive terms when stepping backwards, this making the backward integration stable. If this is not taken into account the model blows up after a few days.

Reversing the diffusion sign in the backward model is a numerical artifact and being so its effects should be carefully analysed. In this section, the backward integration accuracy is studied, as well as its sensitivity with respect to the choice of the diffusion coefficient. The errors are analysed calculating the L2 error norm at the end of one forward-backward integration relative to a typical one day model variation:

$$R_{\text{err}} = \frac{\|\boldsymbol{x}(0) - \tilde{\boldsymbol{x}}(0)\|}{< \|\boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t)\| >} \tag{15}$$

where $\Delta t = 1$day and the brackets represent the empirical mean.

Figure 1 shows the global error, R_{err}, for different window sizes. The errors grow linearly with the window size for all variables. Temperature is the most affected variable, followed by sea level and velocities. Temperature errors exceed 18 times a typical one-day variation for the 30 days experiment and 1.2 times for the 2 days. The use of reduced diffusion/viscosity coefficients reduces the errors to 6.8 and 0.16 times the one-day variation for 30 and 2 days experiments, respectively. Velocities errors were reduced by 50% for 30 days and 85% for 2 days, while ssh errors were reduced by 60% and 88% for 30 and 2 days, respectively.

As shown on Fig. 2 velocity and temperature errors are depth-dependent. Whereas for velocity they are larger at the surface and decrease with depth, for temperature they are larger in the thermocline. In the cases for which the forward-backward integrations use the same diffusion/viscosity coefficients as in the reference simulation, the temperature errors at thermocline depths exceed 3 times the typical one day variation for the 5 days experiments and reaches 15 times for 20 days experiments. Considering the velocities, errors are proportional to 4 one-day variations for the 5 days experiment and to 8 one-day variations for the 20 days experiments. For time windows of 10, 20 and 30 days, velocities at the thermocline depths start to be influenced by temperature errors.

Reduction of the diffusion/viscosity coefficients greatly reduced the errors especially in the thermocline for the temperature and at the surface for the velocity. It can be noted that when the diffusion coefficient is decreased the errors converge to a limit. This limit changes with respect to the window length and should be related to the diffusion required to stabilize the numerical method, which is of second order in our case, and hence oscillatory. Therefore, there is a compromise between the errors

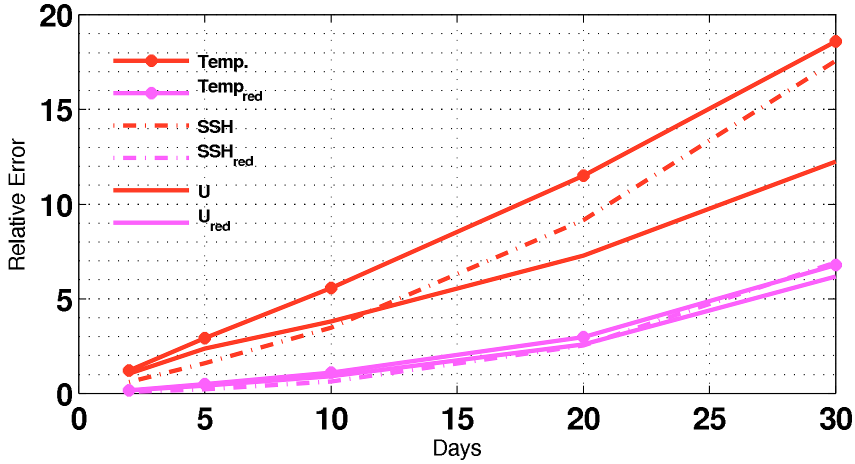

Fig. 1. Errors on the initial condition after one forward-backward model integration perfectly initialized and without nudging. Red curves were obtained using the same diffusion coefficients as in the reference experiment ($\nu_h^{u,v} = -8 \times 10^{10} m^4/s$ and $\nu_h^{t,s} = -4 \times 10^{11} m^4/s$) and magenta curves were obtained using reduced diffusion ($\nu_h^{u,v} = -8 \times 10^{9} m^4/s$ and $\nu_h^{t,s} = -8 \times 10^{10} m^4/s$). The abscissa represents the length of the time window.

induced by the extra diffusion and errors due to spurious oscillations.

Numerical errors were assessed by changing the model time step from $900s$ to $90s$. The resulting errors (not shown) do not change, suggesting that the errors induced by the diffusion are dominant. On the one hand, this is important because the complete rewriting of the model's code can be difficult, similarly to the adjoint model programming used by the 4Dvar, but on the other hand if the assimilation cannot control the diffusion errors it may represent a fundamental problem of the method when it is applied to non-reversible geophysical systems such as the ocean.

Figure 3 shows the spatial structures of the sea level error for the 10 days experiment. The errors are highly variable in space, being larger along the main jet axis. This is probably due to the fact that the backward integration smooths the gradients and so the largest errors are found near the fronts. Therefore, the errors structures may be of high variability in space and time since they are state dependent.

Figure 4 shows the surface kinetic energy spectrum calculated from the experiment employing the reference diffusion coefficient and a reduced diffusion coefficient. The backward integration introduces an extra diffusion, coarsening the effective model resolution, which is defined as the portion of the spectra for which there is a change in the spectrum slope. In the reference simulation the effective model resolution is estimated to be 190km, which is coherent with the $\approx 7 \times \Delta x$ estimation of Skamarock (2004).

The longer the time window the greater the portion of the spectra affected. For the experiment employing the reference diffusion coefficient, the divergence between the true spectra and the spec-

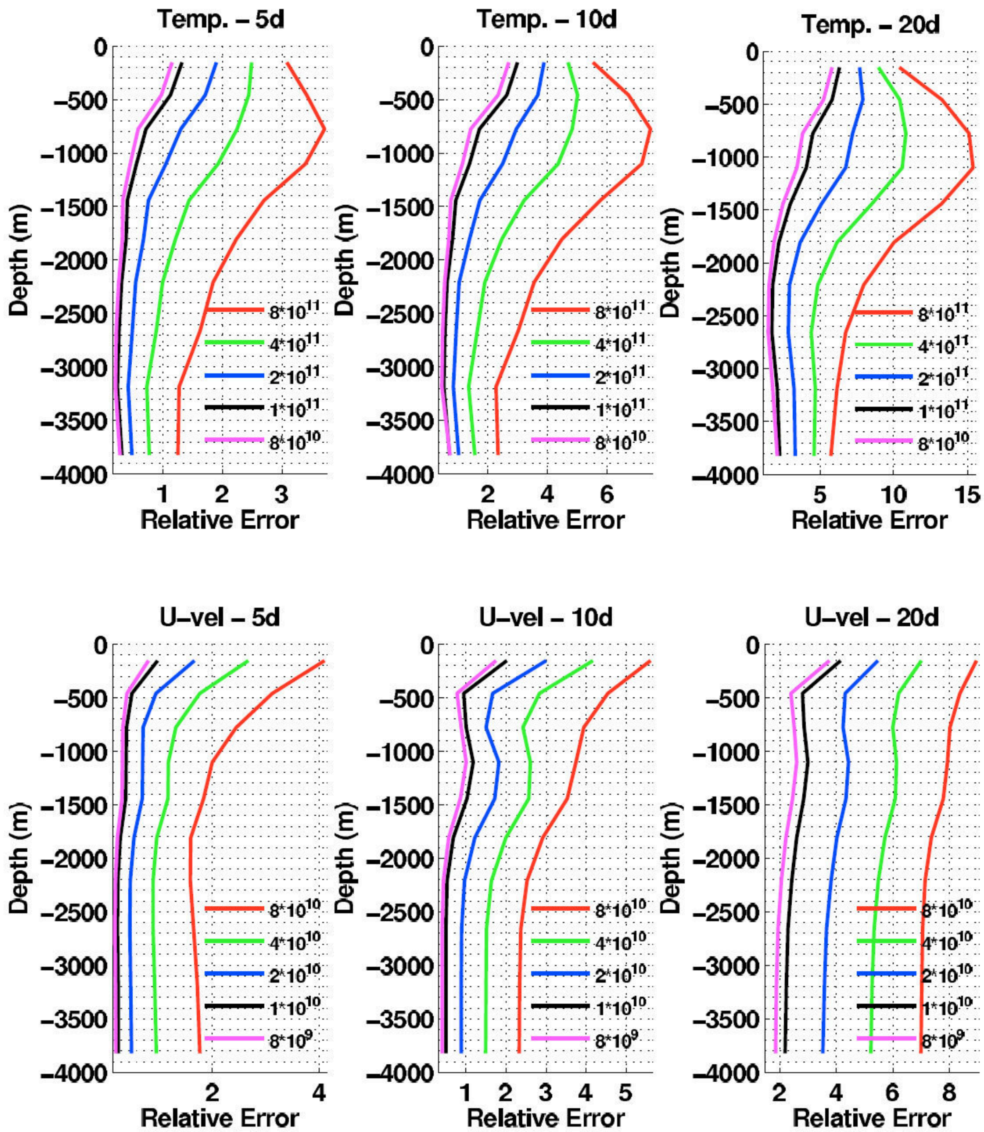

Fig. 2. Vertical profiles of relative errors on the initial condition after one forward-backward model integration without nudging. Each color refers to an experiment performed using the diffusion coefficient indicated in the figures legend. Red curves were obtained using the same diffusion coefficients as in the reference experiment. Top panel: temperature errors; bottom panel: zonal velocity errors. The length of the time window is indicated in the title of each figure.

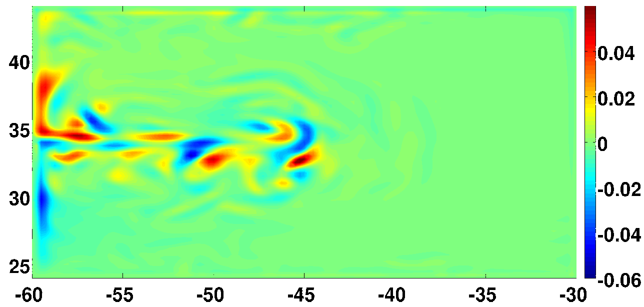

Fig. 3. Sea level errors after one forward-backward model integration. The time window is of 10 days.

tra obtained from the backward integration is observed at 126, 314 and 627km for 5, 10 and 20 days experiments, while for the experiments considering a reduced diffusion coefficient there is almost no differences for the 5 days experiment, and the divergence is observed at 126 and 314km for the 10 and 20 days experiments. If on the one hand using the reduced diffusion helps to keep the energy distribution coherent with the true distribution, on the other hand it creates noise in the range of 126km to 25km. This confirms that there is a trade-off between the errors due to the excessive smoothing and the errors due to high frequency numerical modes.

In this section we have seen that there are large backward-errors induced by over-diffusion. Therefore, short time windows with reduced diffusion coefficients would be preferable to be used in DA experiments. Two regions have to be cautiously analyzed: the surface and the thermocline. Surface layers are prone to feature errors due to their role on the wind energy dissipation while at the thermocline strong density gradients contribute to high diffusion rates.

5 Data Assimilation experiments

5.1 Prescription of the DBFN gain

In this study the increments corresponding to the term $\boldsymbol{K}(\boldsymbol{x}^{obs} - \mathcal{H}(\boldsymbol{x}))$ are calculated in two operations: first the increments of the observed variables are calculated using a prescribed weight and subsequently the increments of the other state variables are calculated using linear regression. More precisely, defining $\boldsymbol{y} = \mathcal{H}(\boldsymbol{x})$ as the observed part of the state vector, the first step may be written as:

$$\delta \boldsymbol{y} = \boldsymbol{\Theta}(\boldsymbol{x}^{obs} - \boldsymbol{y}^{b}) \tag{16}$$

where the superscript ^{b} denotes the background field or the model field available from the last time step. The prescribed weight is given by:

$$\boldsymbol{\Theta} = \frac{\sigma_m^2}{\sigma_m^2 + \gamma \sigma_o^2} \tag{17}$$

where σ_m^2 is the mean spatial value of SSH variance calculated from the free model run, σ_o^2 is the observation error variance and γ is a parameter used to adjust the variance of the observation errors.

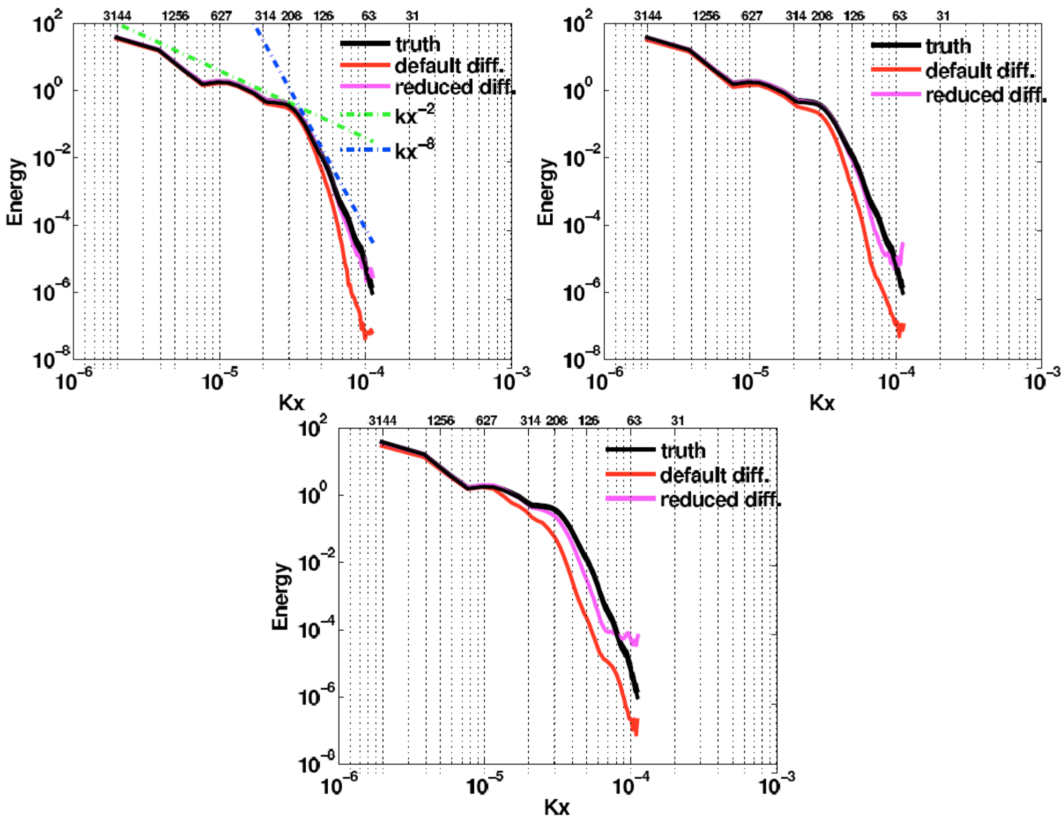

Fig. 4. Kinetic energy mean power spectra calculated using the first layer velocity fields. Black curves represent the "true" initial condition power spectra; Red curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing the reference diffusion coefficient; Magenta curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing a reduced diffusion coefficient. Top left: 5 days assimilation window. Top right: 10 days assimilation window. Bottom: 20 days assimilation window. In the bottom abscissa the ticklabels stand for longitudinal wave-number (rad/m) while in the top abscissa the ticklabels stand for the corresponding wavelengths in km units.

When $\gamma = 1$ the Eq.(17) for the weight Θ has the same form of the scalar Kalman gain (Gelb et al., 1974). For values greater than one, γ is an inflation factor, i.e. it increases the variance of the observation errors resulting in more weight given to the model in the Eq.(16).

The use of the inflation factor is theoretically justified in the linear Kalman filtering context. In this case, it is well-known that the Kalman Filter provides the best linear unbiased estimator. Therefore, there is no need to use more than once the observations. Consequently, when one is iterating the Kalman Filter the inflation parameter should be used to avoide overfitting and the introduction of correlated errors in the system (Kalnay and Yang, 2010). In this study $\gamma = 18$, which means that theoretically the best solution would be reached in 9 iterations. However, since in this study the Kalman Filter equations are not fully used and the system is not linear, the γ parameter is used as a guide on how strong the model is nudge toward the observations. Indeed, the iterations are not limited to 9. The used values for the other parameters are $\sigma_m = 0.017m$ and $\sigma_o = 0.03m$ consistently with the perturbations added to the observations (see Sect 5.4).

Then, the increments of the non-observed variables, $\delta\boldsymbol{x}$, are calculated by using a regression equation of the form:

$$\delta\boldsymbol{x} = \hat{\boldsymbol{B}}^{PLS}\delta\boldsymbol{y} \tag{18}$$

where $\hat{\boldsymbol{B}}^{PLS}$ is the Partial Least Squares (PLS) regression coefficients which are described below. It is worth noting that in Sect 6 we also apply this update scheme to an ordinary direct nudging experiment (ONDG). In this case γ is equal to one.

The PLS can be seen as an improvement to the Ordinary Least Square (OLS) regression. The most important difference between OLS and PLS is that the later assumes that the maximum information about the non-observed variables is in those directions of the observed space which simultaneously have the highest variance and the highest correlation with the non-observed variables.

In the PLS description (Tenenhaus, 1998), $\boldsymbol{Y} \in \mathbb{R}^{n\times M}$ is considered as the observed or predictor variables and $\boldsymbol{X} \in \mathbb{R}^{n\times N}$ as the non-observed or response variables. In our notation n is the sample size and M and N are respectively the size of the state space of $\boldsymbol{Y}$ and $\boldsymbol{X}$. Besides, $\boldsymbol{Y}$ and $\boldsymbol{X}$ are centered and have the same units. The PLS regression features two steps: a dimension reduction step in which the predictors from matrix $\boldsymbol{Y}$ are summarized in a small number of linear combinations called "PLS components". Then, that components are used as predictors in the ordinary least-square regression.

The PLS as well as the principal component regression can be seen as methods to construct a matrix of p mutually orthogonal components $\boldsymbol{t}$ as linear combinations of $\boldsymbol{Y}$:

$$\boldsymbol{T} = \boldsymbol{Y}\boldsymbol{W}, \tag{19}$$

where $\boldsymbol{T} \in \mathbb{R}^{n\times p}$ is the matrix of new components $\boldsymbol{t}_i = (t_{1i},...,t_{ni})^T$, for $i = 1,...,p$, and $\boldsymbol{W} \in \mathbb{R}^{M\times p}$ is a weight matrix satisfying a particular optimality criterion.

The columns $\boldsymbol{w}_1,...,\boldsymbol{w}_p$ of $\boldsymbol{W}$ are calculated according to the following optimization problem:

$$\boldsymbol{w}_i = \text{argmax}_{\boldsymbol{w}}\{cov(\boldsymbol{Y}\boldsymbol{w},\boldsymbol{X})^2\} \tag{20}$$

subject to $\boldsymbol{w}_i^T\boldsymbol{w}_i = 1$ and $\boldsymbol{w}_i^T\boldsymbol{Y}^T\boldsymbol{Y}\boldsymbol{w}_j = 0$ for $j = 1,...,i-1$.

The PLS estimator $\hat{\boldsymbol{B}}^{PLS}$ is given by:

$$\hat{\boldsymbol{B}}^{PLS} = \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{Y}^T\boldsymbol{Y}\boldsymbol{W})^{-1}\boldsymbol{W}^T\boldsymbol{Y}^T\boldsymbol{X} \tag{21}$$

An immediate consequence of Eq. (21) is that when $\boldsymbol{W} = \boldsymbol{I}$ the Ordinary Least Squares solution is obtained.

The number of components p is chosen from cross-validation. This method involves testing a model with objects that were not used to build the model. The data set is divided in two contiguous blocks; one of them is used for training and the other to validate the model. Then the number of components giving the best results in terms of mean residual error and estimator variance is sought.

The weight $\boldsymbol{\Theta}$ and the regression model $\hat{\boldsymbol{B}}^{PLS}$ are kept constant over the assimilation cycles and the correction steps (16) and (17) are applied at the end of the loop of time. Thus, our updating scheme can be seen as a rough approximation of the two steps update for EnKF presented by Anderson (2003).

5.2 The 4Dvar background term configuration

The 4Dvar considers a background term of the form:

$$J_b = \frac{1}{2}(\delta\boldsymbol{x}_0^k)^T\boldsymbol{B}^{-1}(\delta\boldsymbol{x}_0^k)$$

where $\boldsymbol{B}$ is the background error covariance matrix. This term is also known as a regularization term in the sense of Tikhonov. It is specially important when there is not enough observation to determine the problem.

The $\boldsymbol{B}$ matrix is supposed to model the spatial covariance of the background errors of a given variable as well as the cross-covariance between the errors of different variables. Since the state space is too big, it is impossible to store the entire covariance matrix. Therefore, Derber and Bouttier (1999) have proposed the decomposition of the multivariate problem into a sequence of several univariate problems. This is accomplished by decomposing the variables into a balanced component and an unbalanced component. This is done to all variables but one should be kept without decomposition so as we can define the balanced and unbalanced components of the other variables. We used the decomposition proposed by Weaver et al. (2005) for which the temperature is the "seed" variable and then thanks to some physical constraints such as the geostrophic balance, the hydrostatic balance and the principle of water mass conservation all other state variables may be decomposed into a balanced (B) component and an unbalanced (U) component. Thus, each model variable, namely temperature ($\boldsymbol{temp}$), salinity ($\boldsymbol{salt}$), sea surface height ($\boldsymbol{\eta}$), zonal velocity ($\boldsymbol{u}$) and meridional velocity ($\boldsymbol{v}$), may be written as:

$$\boldsymbol{temp} = \boldsymbol{temp} \tag{22}$$

$$\boldsymbol{salt} = \boldsymbol{salt}_B + \boldsymbol{salt}_U = \boldsymbol{G}_{\boldsymbol{salt},\boldsymbol{temp}}(\boldsymbol{temp}) + \boldsymbol{salt}_U \tag{23}$$

$$\boldsymbol{\eta} = \boldsymbol{\eta}_B + \boldsymbol{\eta}_U = \boldsymbol{G}_{\boldsymbol{\eta},\boldsymbol{\rho}}(\boldsymbol{\rho}) + \boldsymbol{\eta}_U \tag{24}$$

$$\boldsymbol{u} = \boldsymbol{u}_B + \boldsymbol{u}_U = \boldsymbol{G}_{\boldsymbol{u},\boldsymbol{\rho}}(\boldsymbol{\rho}) + \boldsymbol{u}_U \tag{25}$$

$$\boldsymbol{v} = \boldsymbol{v}_B + \boldsymbol{v}_U = \boldsymbol{G}_{\boldsymbol{v},\boldsymbol{\rho}}(\boldsymbol{\rho}) + \boldsymbol{v}_U \tag{26}$$

$$\tag{27}$$

where

$$\boldsymbol{\rho} = \boldsymbol{G}_{\boldsymbol{\rho},\boldsymbol{temp}}(\boldsymbol{temp}) + \boldsymbol{G}_{\boldsymbol{\rho},\boldsymbol{salt}}(\boldsymbol{salt}) \tag{28}$$

$$\boldsymbol{p} = \boldsymbol{G}_{\boldsymbol{p},\boldsymbol{p}}(\boldsymbol{\rho}) + \boldsymbol{G}_{\boldsymbol{p},\boldsymbol{\eta}}(\boldsymbol{\eta}) \tag{29}$$

with $\boldsymbol{\rho}$ the density and $\boldsymbol{p}$ the pressure.

Then, since a covariance matrix may be written as the product of variances and correlations, $\boldsymbol{B}$ may be expressed as:

$$\boldsymbol{B} = \boldsymbol{G}\boldsymbol{\Lambda}^T\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{G}^T$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of error standard deviation, for which the climatological standard deviation are the entries, and $\boldsymbol{C}$ is an univariate correlation matrix modeled using the generalized diffusion equation (Weaver and Courtier, 2001; Weaver et al., 2005). In this method the user should chose typical decorrelation lengths. In this study the horizontal decorrelation length is set to $400km$ and the vertical decorrelation length is set to $1500m$. In addition, the 4Dvar is configured to perform one outer-loop and a maximum of thirty inner-loop for each assimilation cycle.

5.3 Assimilation cycle

One assimilation cycle is defined as the process of identifying an initial condition through the iterative process followed by a forecast spanning the assimilation window, which provides a new background to the next assimilation cycle.

The objective of cycling is to provide a background state for the next assimilation window that is closer to the true state than the very first background field. This usually reduces the number of iterations needed by the algorithms to reach convergence.

The length of the Data Assimilation window (DAw) used in the reference experiments (Sect 6.1) is 10 days. For the sensitivity experiments presented in the Sect 6.2 the lengths of the the assimilation window are 5 days and 30 days.

5.4 Observation network

In this article, every four days an observation network simulating Jason-1 satellite density sample is available. The data is perturbed with white Gaussian noise with standard deviation equals to $3cm$. With this observation network a new set of 5000 observations is available every four days.

The data assimilation problem we proposed to solve is to recover the full model state at the beginning of the assimilation window. The model state space is composed of five variables: sea surface height ($\boldsymbol{\eta}$), meridional and zonal velocities ($\boldsymbol{u}$ and $\boldsymbol{v}$), temperature and salinity ($\boldsymbol{temp}$ and $\boldsymbol{salt}$). Since we have a horizontal mesh of size 81 x 121 and 11 vertical layers the total size of the state space is 441045. Therefore, the problem is undetermined, since the observations represent only a 1.1% of the total state space. This means that the background term, and accordingly the $\boldsymbol{B}$ matrix for the 4Dvar and the regression model $\hat{\boldsymbol{B}}^{PLS}$ for the DBFN, have quite a strong importance on the method performances since they project the increments of the observed variables onto the numerous non-observed variables.

To study at which extent the results are depend on the amount of assimilated observations and on the first guess, in Sect 6.2.2 two additional experiments assimilating complete daily fields of SSH are conducted: one using the same first guess of the experiments of Sect 6.1, and another using a perturbed initial condition. In despite of the fact that the problem continues to be underestimated, in this case the SSH analysis is no more dependent on the SSH spatial covariance, and the unstable modes associated with the SSH dynamics are certainly observed. The analysis produced for the other state vector variables remains dependent on the matrices $\mathbf{B}$ for the 4Dvar case and $\hat{B}^{PLS}$ for the DBFN case.

6 Data Assimilation Results

6.1 Reference experiment

In this section the results produced by the DBFN, the 4Dvar method, the Ordinary Nudging (ONDG) and the control experiment are presented. All assimilation methods include the five prognostic variables in the state vector. This is possible thanks to the PLS regression method in the case of the DBFN and ONDG and thanks to the multivariate balance operator $\boldsymbol{G}$ in the case of the 4Dvar experiments. The diffusion and viscosity coefficients used in the DBFN experiments are those which produced the smaller errors in the experiments without Nudging, as reported in Sect 4.

First the minimization performance of the 4Dvar implementation is analysed. Figure 5 shows the reduction of the cost function gradient for the 4Dvar and the reduction of the relative error of the zonal velocity for the DBFN, both of them for the first assimilation cycle. 4Dvar takes 26 iterations to approximately achieve the optimality condition $\nabla J = 0$. This represents 3 times the number of iterations required by the DBFN to converge, i.e., after which the errors cease to decrease. Moreover,

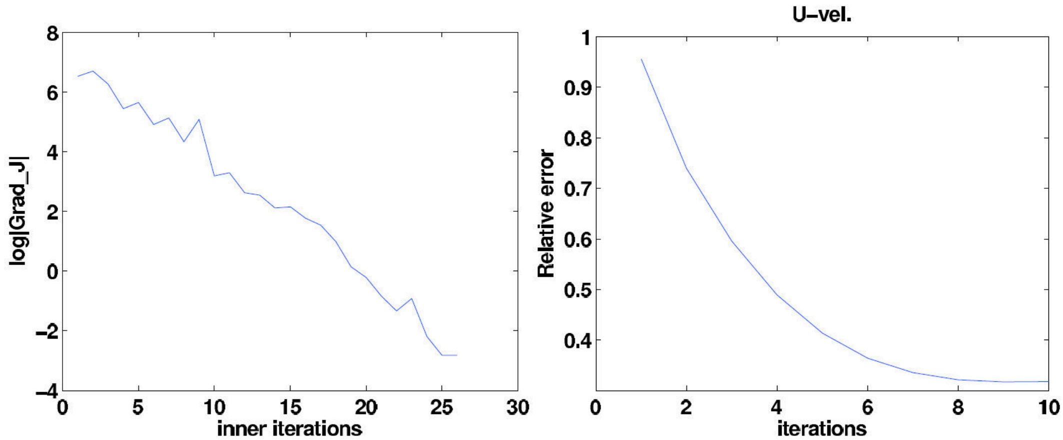

Fig. 5. Figure shows the gradient of the cost function after each inner iteration (left) and the reduction of the relative error for zonal velocity for the DBFN experiment (right).

the 4Dvar numerical cost is more than 3 times the DBFN cost since one execution of the adjoint model costs four times the cost of the direct model in terms of CPU time.

We note that the minimum error for the DBFN is reached after 9 iterations. This is quite consistent with our choice $\gamma = 18$, since theoretically it allows the use of the same set of observations for 18 times.

At this point we find appropriate to present the fact that the trajectories of the forward and backward nudging are very close to each other at convergence, which justifies the qualitative explanation of the DBFN algorithm given by Eqs. (6) and (7). This fact can be seen in the Fig 6 that shows the forward and backward mean surface zonal velocity trajectories at convergence as well as the surface zonal velocity trajectories for a point located on the unstable jet, at 34^o North and 52.6^o West.

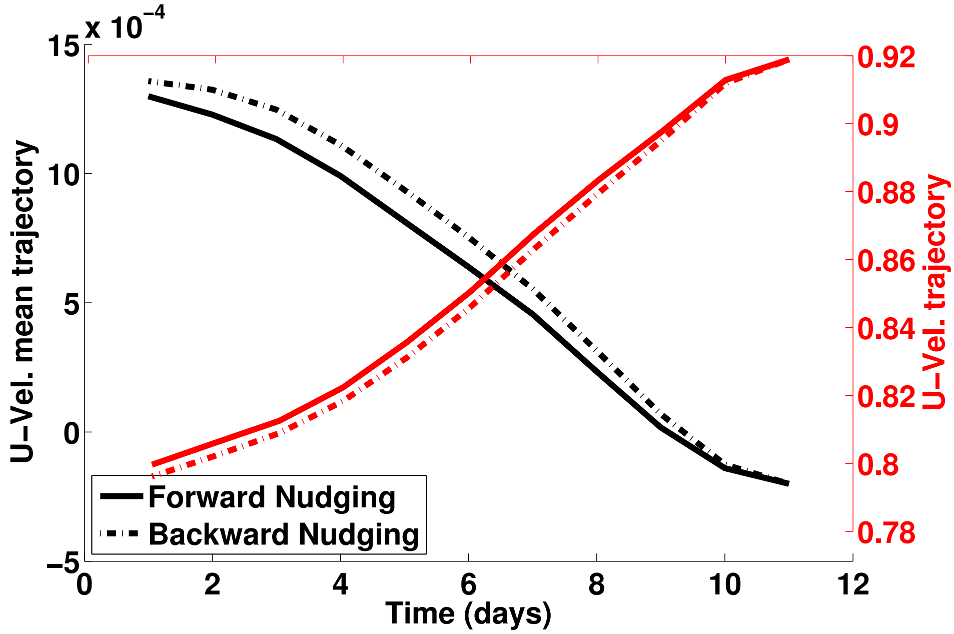

Fig. 6. Black curves represent the forward and backward mean surface zonal velocity trajectories at convergence and red curves the forward and backward surface zonal velocity trajectories at convergence for a point located at 34^o North and 52.6^o West, which is located on the unstable jet.

Figure 7 shows the root mean squared (rms) error for the control experiment (without assimilation), the experiment using the direct nudging with PLS regression (ONDG), the DBFN and the 4Dvar. The DBFN errors for the velocity and SSH converge to their asymptotic values after the first assimilation cycle while for ONDG and 4Dvar errors stop decreasing after 100 and 200 days, respectively. This is a benefit of the iterations performed by the DBFN when model and data are quite different. Among the experiments conducted, the DBFN produced the smallest errors for all variables, except for the zonal velocity, for which the 4Dvar has slightly smaller errors. The ONDG also showed good performance, but with errors larger than the DBFN and 4Dvar errors.

With respect to the vertical error (Fig. 8), the DBFN and the ONDG performed better for the upper ocean than 4Dvar. Clearly, the PLS also corrects the deep ocean velocity, but less accurately than 4Dvar. The first error mode is the barotropic one, i.e. it has the same sign over all depths, and accounts for 97% of the error variability for 4Dvar, 96% and 93% for DBFN and ONDG, respectively. Although the first mode is the barotropic one for all methods, the 4Dvar barotropic mode of error is out of phase with respect to the PLS barotropic mode. This reflects the better performance of the 4Dvar for the deep ocean and the better performance of the DBFN and ONDG for the upper ocean.

The second mode, which accounts for almost all the remaining variability, has a sign inversion with depth and is found especially over the main axis of the jet. In this region the deep ocean veloc-

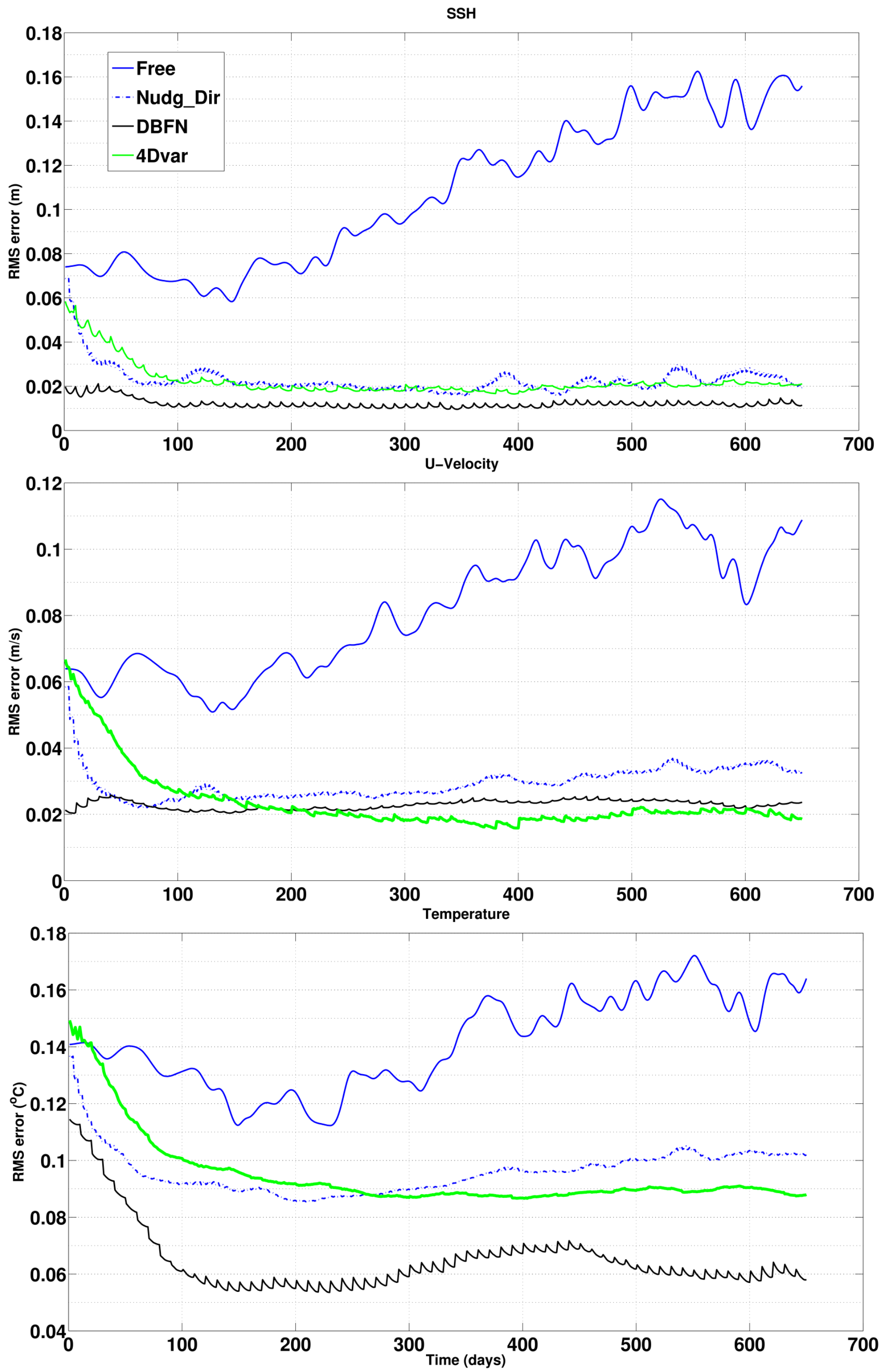

Fig. 7. The figure shows errors of the SSH (top panel), the zonal velocity (middle panel) and the temperature (bottom panel).

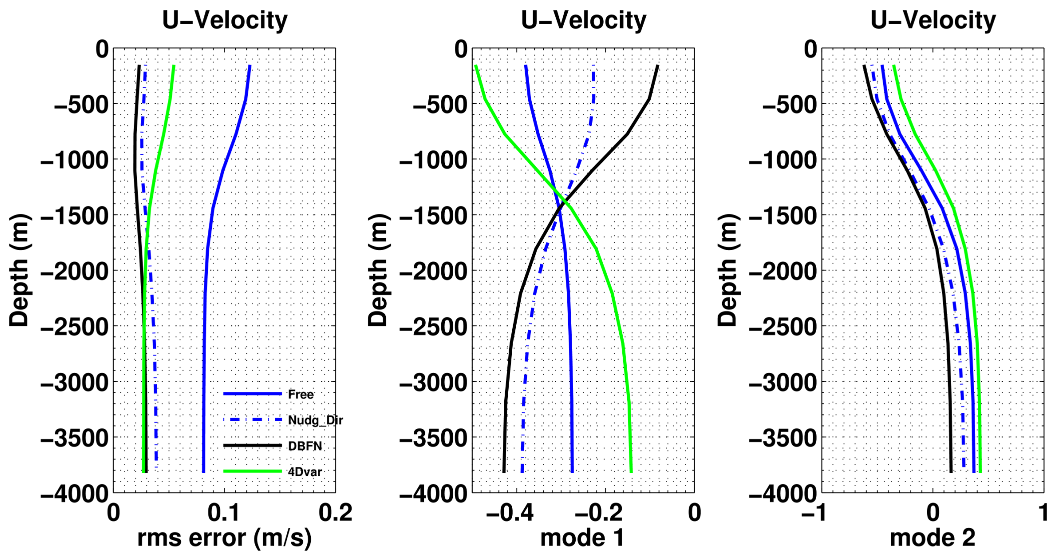

Fig. 8. Vertical profiles of rms error in zonal velocity (Left panel) and first (middle panel) and second (right panel) eof error modes calculated using forecast from day 200 to day 720.

ities are overestimated due to spurious covariances between the SSH and the deep ocean velocities.

The way both methods correct the model depends on the $\boldsymbol{B}$ matrix in the 4Dvar algorithm and on the regression model $\hat{B}^{PLS}$ in the DBFN. It means that results may be different if another approximation of $\boldsymbol{B}$ and another model regression model are used. Perhaps the main conclusion of this comparison is that the DBFN, which is easier to implement and cheaper to execute, can produce results similar to 4Dvar. Also, it is shown that iterations is an important aspect of the method. Iterations compensate for the lack of a priori information on the model background errors as well as filter out noise in observations. The latter must be connected to the diffusive character of the algorithm. Moreover, the iterations allows us to put information from the observations into the model, without causing initialization problems since the nudging gain can be taken smaller than the one used for the direct nudging due to the possibility of using more than once the same set of observations.

6.2 Sensitivity experiments

6.2.1 Length of the DAw

Sensitivity tests with respect to the length of the DAw are presented. As we have shown in Sect 4, the accuracy of the backward model is inversely proportional to the length of the DAw. Therefore, in this section we present experiments using a DAw of five days and thirty days. The experiments configuration is similar to those presented in the previous section.

Figure 9 shows the evolution of the rms errors for the zonal velocity and temperature during the

DBFN iterations over the first assimilation cycle, for three DAw (including the ten day-window used previously). When considering only one iteration, the best results were obtained with the 30 days-window experiment. This is a consequence of the asymptotic character of the Nudging method: the longer the assimilation window, the more observations accounted for, the smaller the error. This changes when several iterations are considered. The observed divergence for the 30 days-window is due to the errors induced by the over-diffusion that induce great increments, which by their nature, are not modelled by the ensemble of model states used to construct the regression model.

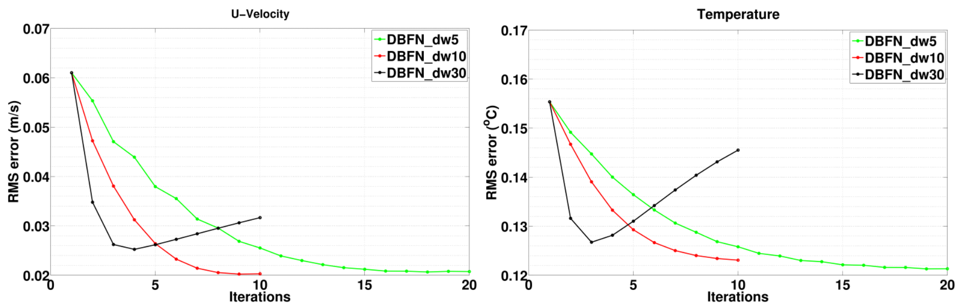

Fig. 9. Evolution of the rms errors for the zonal velocity and temperature during the DBFN iterations over the first assimilation cycle, for three DAw: 5, 10 and 30 days.

Figure 10 shows the rms error for the DBFN and 4Dvar experiments for three DAw: 5, 10 and 30 days. The methods exhibited comparable performance depending on the length of the DAw. For the DBFN the 5 and 10 days DAw provided better results than the 30 days window, while for the 4Dvar the 30 days window provided the best estimation in terms of rms error. The DBFN and 4Dvar experiments using the 30 and 5 days DAw, respectively, failed to identify the initial conditions since their SSH rms errors are greater than the observation error standard deviation.

Figure 11 shows the time evolution of vertical profiles of horizontally layer-wise averaged rms error of zonal velocities for the DBFN and 4Dvar experiments. The 4Dvar profits of the longer DAw to spread the observation to the 3-dimensional variables. This is done by the iterations of the direct model and by the $\boldsymbol{B}$ matrix. For the DBFN experiments, after one year of data assimilation the errors in the deep ocean start to grow. This is due to the high variance of the PLS estimator for deep layers. The problem becomes more evident on the second year because at this stage the observations are farther from the model states used to construct the regression coefficients. Therefore, this mean that this behavior is not intrinsic to the DBFN algorithm and its diffusive aspects, but due to our implementation. Ideally, the regression model should evolve in time, similarly to the Kalman Filter scheme. The 4Dvar has good performance at the deep ocean thanks to the use of a vertical localization with a length scale of $1500m$.

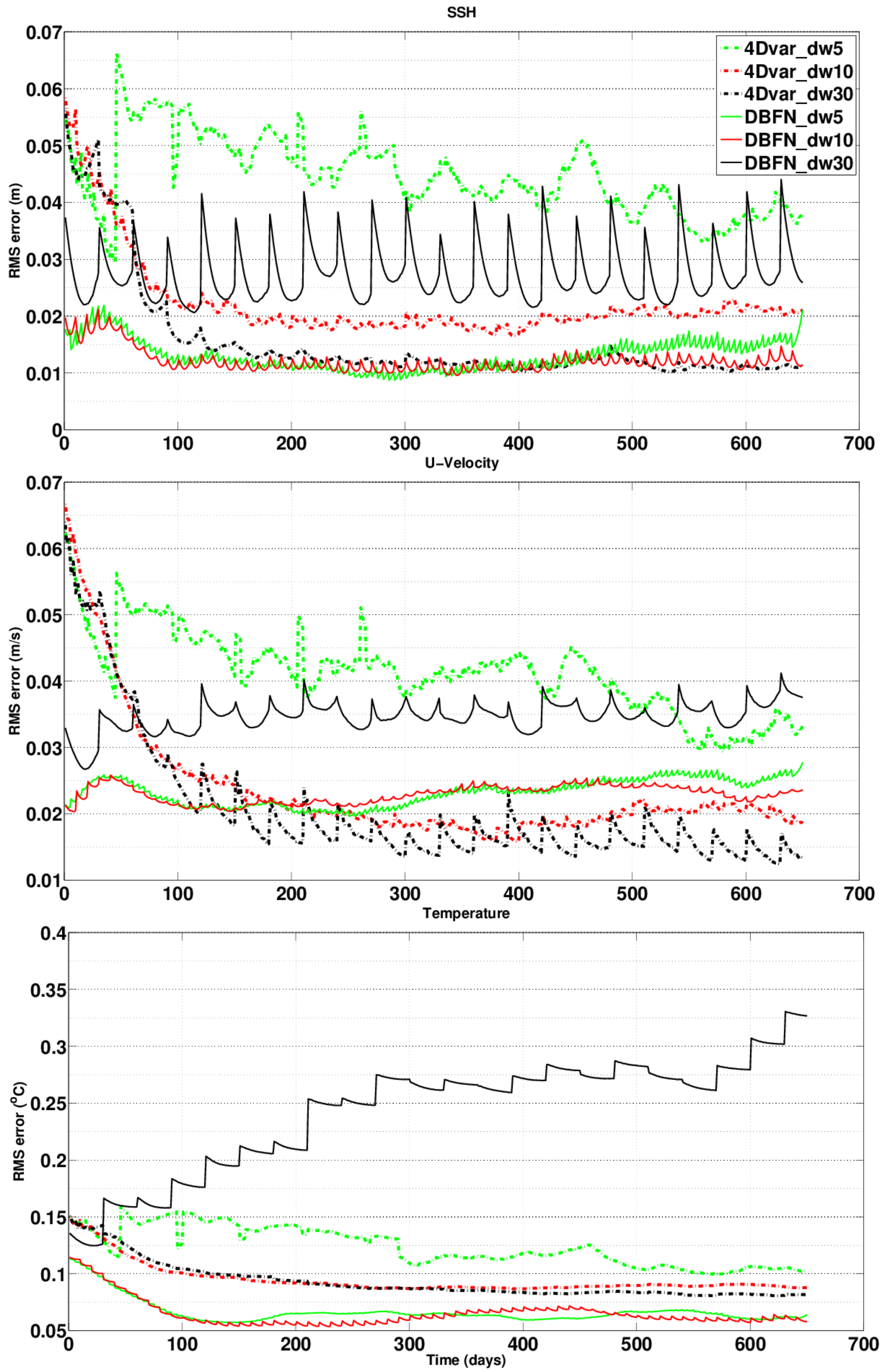

Fig. 10. RMS errors on SSH (top panel), zonal velocity (middle panel) and temperature (bottom panel) from DBFN and 4Dvar experiments with DAw of 5, 10 and 30 days.

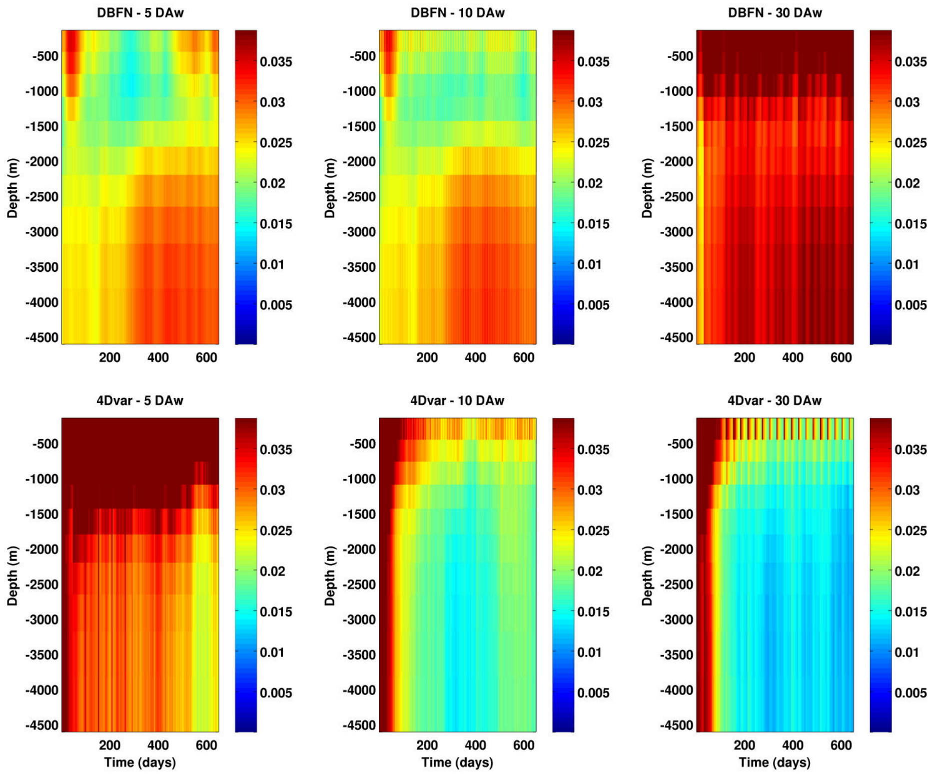

Fig. 11. Time evolution of vertical profiles of horizontally layer-wise averaged rms error of zonal velocities for the DBFN (top panels) and 4Dvar (bottom panels) experiments. Units are in (m/s).

Next we investigate which scales are better represented by each assimilation method. This is done by comparing the surface kinetic energy spectrum and the deep ocean kinetic spectrum produced by each method. The Fig.(12) shows that the effective resolution of the model is not affected by the diffusive character of the DBFN algorithm. It is clear that there is a reduction of the energy for the scales close to the grid scale, but the energy contained in scales greater than $7 \times \Delta x$ is not affected. It means that the diffusion-induced errors presented in Sect 4 are ”controlled“ by the assimilation of sea surface height observations.

There is no great difference between the DBFN and 4Dvar surface spectrum for the assimilation windows shorter than 30 days, which once more proves the reliability of the DBFN for the assimilation of oceanic observations. The energy spectra for the deep ocean velocities produced by the DBFN contains more energy than the true spectrum independently of the used DAw. This confirms that the deep ocean velocity errors are due to the high variance of the PLS regression model.

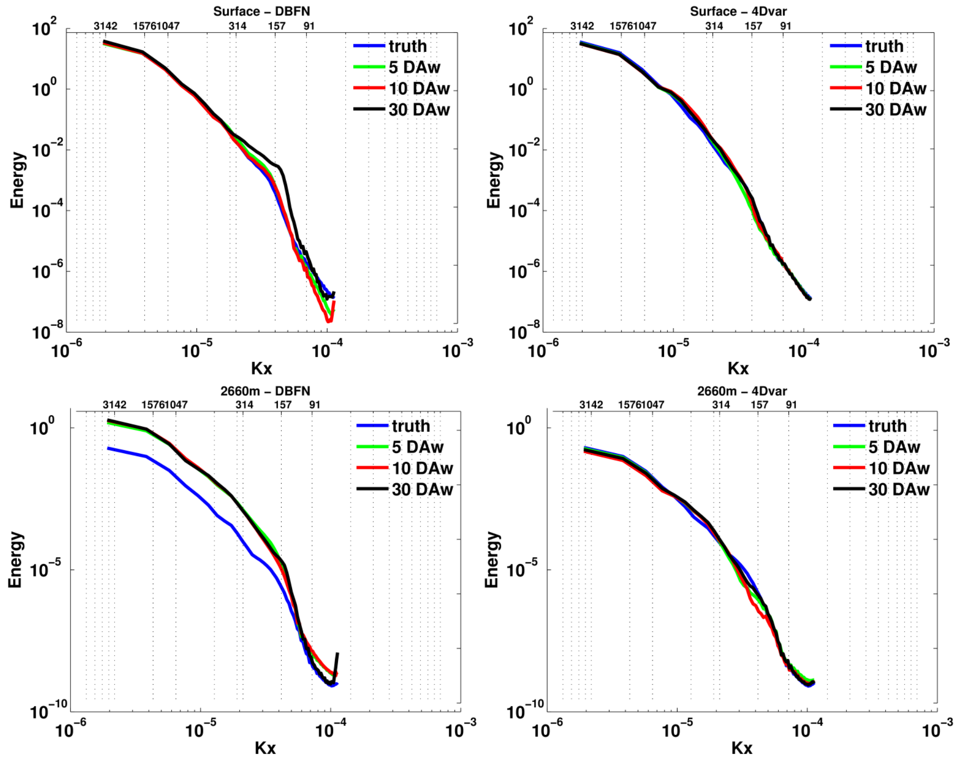

Fig. 12. Kinetic energy mean power spectra calculated using the first layer (top) and a layer at 2660m (bottom) and using the 650 days of the assimilation experiments using the DBFN (left) and the 4Dvar (right). Blue curves represent the "true" power spectra; Green curves represent the power spectra calculated for the 5 days DAw; Red curves represent the power spectra calculated for the 10 days DAw and Black curves represent the power spectra calculated for the 30 days DAw. In the bottom abscissa the tick-labels stand for longitudinal wave-number (rad/m) while in the top abscissa the tick-labels stand for the corresponding wavelengths in km units.

6.2.2 Observations density and first guess

Finally, two new experiments similar to the one presented in the Sect 6.1 and assimilating complete daily fields of SSH are presented. The first one uses the same initial condition of the previously presented experiments and its goal is to investigate the role of the amount of assimilated observations on the results. In despite of the fact that the problem continues to be underestimated, in this case the SSH analysis is no more dependent on the SSH spatial covariance, and the unstable modes associated with the SSH dynamics are certainly observed. The analysis produced for the other state vector variables remains dependent on the matrices $\mathbf{B}$ for the 4Dvar case and $\hat{B}^{PLS}$ for the DBFN case.

Fig.13 shows the rms error for the SSH and zonal velocity. The results are quite similar to the results presented in Sect 6.1 with a lower rms error for all variables for both methods. Fig.14 shows the initial condition error for the zonal velocity produced by both methods for the satelite-like observations and the complete observations experiments. The figure reveals that while in some places the inclusion of more observations helps to reduce the error in other places it increases the error. This means that although much more information could be extracted from the new set of observations, which decreases the global rms errors, the solution remains dependent on the covariance structures contained on $\mathbf{B}$ and $\hat{B}^{PLS}$.

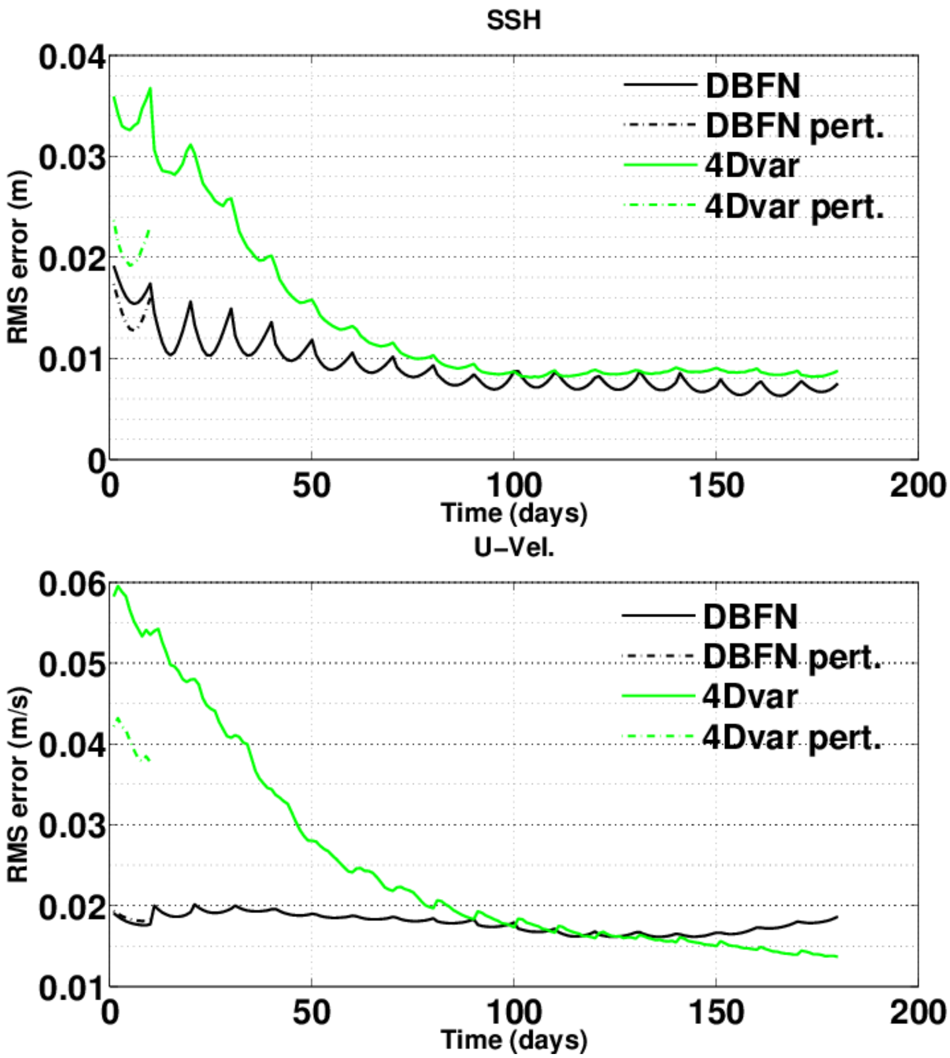

Fig. 13. RMS errors of SSH (top panel) and zonal velocity (bottom panel) from the DBFN and 4Dvar experiments with DAw of 10 days and assimilating complete daily fields of SSH. Dashed lines concern the results from the perturbed experiments.

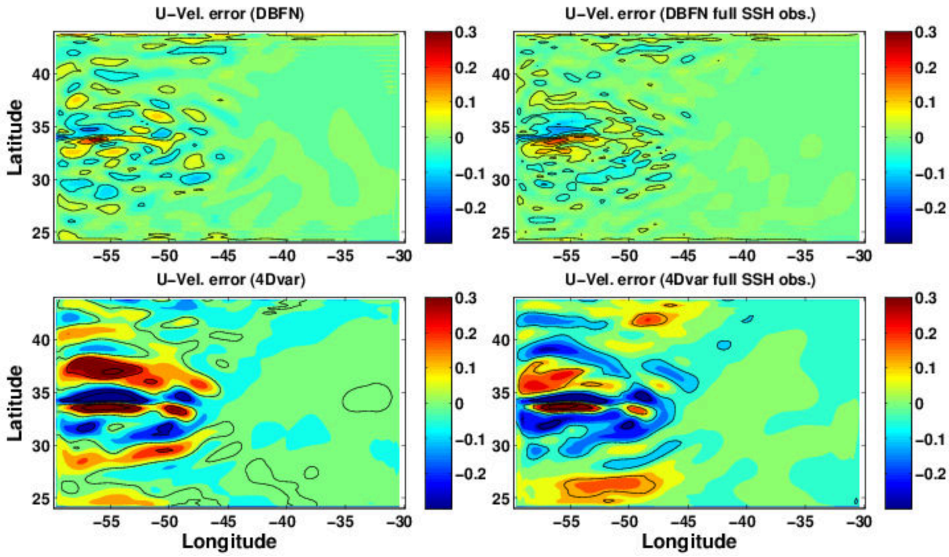

Fig. 14. Zonal velocity error (analysis - truth) for the first assimilation cycle from DBFN experiments (top panels) and 4Dvar experiments (bottom panels). Right panels show the results obtained by assimilating complete daily fields of SSH and the left panels the results from the experiments presented in the Sect 6.1.

The second experiment is initialized with an initial condition that is 20 days apart from the one used previously, and is closer in terms of rms error to the observations. We call this experiment as perturbed experiment. In this case, the objective is to analyze the sensitivity of the solution to the choice of the first guess. Thus, only one assimilation cycle is performed.

Fig.15 shows the initial condition error for the SSH produced by both methods for the perturbed and non-perturbed experiments. Since the perturbed initial condition is not much different from the unperturbed one, the analysis errors have the same structure in both cases, but they differ from one method to another.

The DBFN produced smaller differences between the perturbed and non-perturbed experiences than the 4Dvar for the entire domain. A remarkable difference between the errors produced by the 4Dvar and the DBFN is the error structure in the western boundary that is produced by the DBFN, which is positive northward $34^o N$ and negative southward $34^o N$. The presence of this structure is related to the fact that the DBFN analysis is the final condition produced by the backward model. The same pattern was also observed in the Fig. 3 that shows the backward error for the SSH variable. The rms error of the identified trajectory for the perturbed experiment may be seen in Fig. 13 as the green (4Dvar) and black (DBFN) dashed curves. The results clearly show that for the configured experiments the DBFN is much less sensitive to the first guess than the 4Dvar.

The small sensitivity of the DBFN to the first guess is in accordance with the theoretical result about the BFN presented by Auroux and Blum (2005) that states that for a linear system and under complete observation condition the identified trajectory is independent of the first guess. To what extent this theoretical result may be extended to nonlinear systems assimilating an incomplete set of observations, as the one studied in this article, we do not know. The results presented here suggest that the use of the DBFN may be advantageous in situations in which the system passes by strong changes resulting in a background (first guess) that is far from the true state.

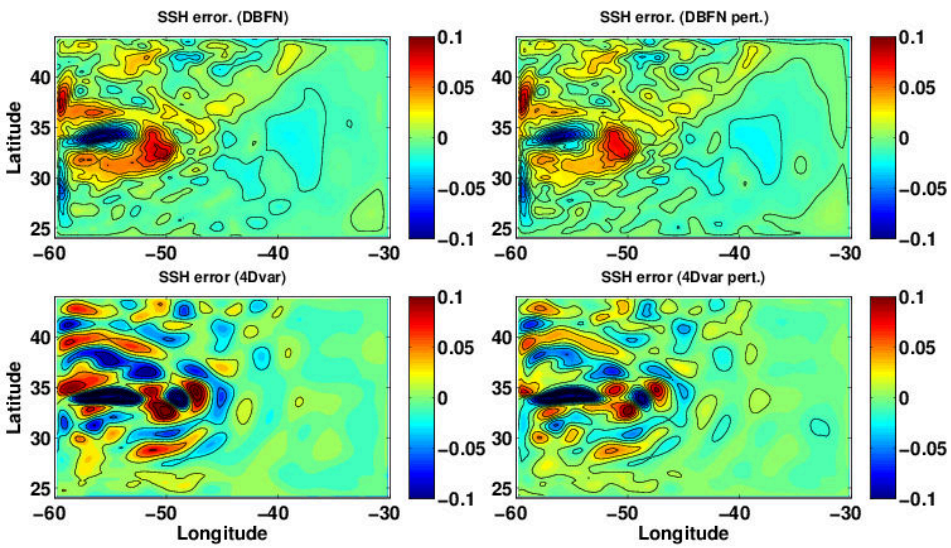

Fig. 15. SSH error (analysis - truth) from DBFN experiments (top panels) and 4Dvar experiments (bottom panels). Right panels show the results obtained from the perturbed experiment.

7 Conclusions and perspectives

This study used the NEMO general circulation model in a double gyre configuration to investigate the Diffusive Back and Forth Nudging performance under different configurations of the data assimilation window, observation network and initial conditions, and to compare it with 4Dvar.

It has been shown that the reliability of the backward integration should be carefully examined when the BFN/DBFN is applied to non-reversible systems. This should support the choice of the assimilation window and identify whether the available observations are sufficient to control the errors induced by the non-reversible terms of the model equations. In this article we have shown that the DBFN might be used for the assimilation of realistically distributed ocean observations, despite the limited accuracy of the backward integration. Improving the backward integration would further improve the DBFN performance and make possible the use of longer assimilation windows.

Our results show that the DBFN can produce results comparable to 4Dvar using lower computational power. This is because DBFN demands less iterations to converge and because one iteration of 4Dvar corresponds to one integration of the tangent linear model, one integration of the adjoint model, which costs four times more than one standard model integration, plus the cost of minimizing the cost function, while the DBFN costs twice the integration of the nonlinear model.

The sensitivity tests show that for the 4Dvar long assimilation windows should be preferably used because it favors the propagation of the sea surface height information to the deep layers. For the DBFN, short windows are preferable because it reduces the effect of the diffusion-induced errors. In future works it would be beneficial to account for this errors when constructing the nudging gain.

Moreover, the results show that for assimilation systems assimilating a much reduced number of observations with respect to the size of the state space, such as ocean data assimilation systems usually do, the set-up of the covariance matrix is a key step since this matrix propagates the information from the observed variables to the non-observed variables. In addition, although in this study the methods have been configured with different covariance matrices, the results show that the DBFN is less sensitive to the background field than the 4Dvar.

Finally, it appears that the DBFN algorithm is worth being further explored both on theoretical and practical aspects, especially those related to the optimization of the matrix $\boldsymbol{K}$ and applications to a more realistic configuration.

Acknowledgements. This work was supported by CNRS/INSU through the LEFE/MANU program. This work was granted access to the HPC and visualization resources of "Centre de Calcul Interactif" hosted by "Université Nice Sophia Antipolis". Calculations were also performed at the IDRIS computational facility center. The authors thanks Pierre-Antoine Bouttier for his help with the set-up of the 4Dvar algorithm. We also thank the Editor and the reviewers for their interesting comments that have enabled the improvement of the manuscript.

References

Abarbanel, H. D. I., Kostuk, M., and Whartenby, W.: Data assimilation with regularized nonlinear instabilities, Quarterly Journal of the Royal Meteorological Society, 136, 769–783, doi:10.1002/qj.600, http://dx.doi.org/10.1002/qj.600, 2010.

Anderson, J. L.: A local least squares framework for ensemble filtering, Monthly Weather Review, 131, 634–642, 2003.

Anthes, R. A.: Data Assimilation and Initialization of Hurricane Prediction Models, J. Atmos. Sci., 31, 702–719, doi:10.1175/1520-0469(1974)031⟨0702:DAAIOH⟩2.0.CO;2, http://dx.doi.org/10.1175/1520-0469(1974)031⟨0702:DAAIOH⟩2.0.CO;2, 1974.

Auroux, D.: The back and forth nudging algorithm applied to a shallow water model, comparison and hybridization with the 4D-VAR, Int. J. Numer. Methods Fluids, 61, 911–929, 2009.

Auroux, D. and Blum, J.: Back and forth nudging algorithm for data assimilation problems, C. R. Acad. Sci. Paris, Ser. I, 340, 873–878, 2005.

Auroux, D. and Blum, J.: A nudging-based data assimilation method for oceanographic problems: the Back and Forth Nudging (BFN) algorithm, Nonlin. Proc. Geophys., 15, 305–319, 2008.

Auroux, D. and Nodet, M.: The back and forth nudging algorithm for data assimilation problems: theoretical results on transport equations, ESAIM Control Optim. Calc. Var., 18, 318–342, 2012.

Auroux, D., Blum, J., and Nodet, M.: Diffusive Back and Forth Nudging algorithm for data assimilation, C. R. Acad. Sci. Paris, Ser. I, 349, 849854, 2011.

Auroux, D., Bansart, P., and Blum, J.: An evolution of the Back and Forth Nudging for geophysical data assimilation: application to Burgers equation and comparisons, Inv. Prob. Sci. Eng., 21, 399–419, 2012.

Ballabrera-Poy, J., Kalnay, E., and Yang, S.-C.: Data assimilation in a system with two scalescombining two initialization techniques, Tellus A, 61, 539–549, doi:10.1111/j.1600-0870.2009.00400.x, 2009.

Bergemann, K. and Reich, S.: A mollified ensemble Kalman filter, Quarterly Journal of the Royal Meteorological Society, 136, 1636–1643, doi:10.1002/qj.672, http://dx.doi.org/10.1002/qj.672, 2010.

Blayo, E., Verron, J., and Molines, J.-M.: Assimilation of topex/poseidon altimeter data into a circulation model of the north atlantic., Journal of Geophysical Research: Oceans, 99, 24 691–24 705, 1994.

Blum, J., Le Dimet, F.-X., and Navon, I. M.: Data Assimilation for Geophysical Fluids, in: Computational Methods for the Atmosphere and the Oceans, edited by Ciarlet, P. G., Temam, R., and Tribbia, J., vol. 14 of *Handbook of Numerical Analysis*, pp. 385–442, Elsevier, Oxford, United Kingdom, 2008.

Bonnans, F. and Rouchon, P.: Commande et optimisation de systèmes dynamiques, Les Editions de l'École polytechnique, 2005.

Bouttier, P.-A., Blayo, E., Brankart, J. M., Brasseur, P., Cosme, E., Verron, J., and Vidard, A.: Toward a data assimilation system for NEMO, Mercator Ocean Quarterly Newsletter, 46, 31–45, 2012.

Chang, K. I., Ghil, M., Ide, K., and Lai, C. C. A.: Transition to aperiodic variability in a wind-driven double-gyre circulationmodel., J. Phys. Oceanography, 31, 1260–1286, 2001.

Chassignet, E. P. and Gent, P. R.: The influence of Boundary Conditions on Midlatitude Jet Separation in Ocean Numerical Models., J. Phys. Oceanography, 21, 1290–1299, 1991.

Chen, X., Liu, C., ODriscoll, K., Mayer, B., Su, J., and Pohlmann, T.: On the nudging terms at open boundaries in regional ocean models, Ocean Modelling, 66, 14 – 25, doi:10.1016/j.ocemod.2013.02.006, http://www.

sciencedirect.com/science/article/pii/S1463500313000401, 2013.

Clifford, M., Horton, C., Schmitz, J., and Kantha, L. H.: An oceanographic nowcast/forecast system for the Red Sea, Journal of Geophysical Research: Oceans, 102, 25 101–25 122, doi:10.1029/97JC01919, http://dx.doi.org/10.1029/97JC01919, 1997.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a Reduced-rank, square-root smoother for ocean data assimilation, Ocean Modelling, 33, 87–100, 2010.

Courtier, P., Thepaut, J. N., and Hollingsworth, A.: A strategy for operational implementation of 4d-var, using an incremental approach., Q. J. R. Meteorol. Soc., 123, 1367–1387, 1994.

Donovan, A., Mirrahimi, M., and Rouchon, P.: Back and forth nudging for quantum state reconstruction, in: 4th International Symposium on Communications, Control and Signal Processing, pp. 1–5, Limassol, Cyprus, 2010.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, doi:10.1029/94JC00572, http://dx.doi.org/10.1029/94JC00572, 1994.

Gelb, A., Kasper, J., Nash, R. A., Price, C. F., and Sutherland, A. A.: Applied Optimal Estimation, The M.I.T. Press, Reading, Massachusetts, arthur gelb edn., 1974.

Haines, K., Malanotte-Rizzoli, P., Young, R. E., and Holland, W. R.: A comparison of two methods for the assimilation of altimeter data into a shallow-water model, Dynamics of Atmospheres and Oceans, 17, 89 – 133, doi:10.1016/0377-0265(93)90014-X, http://www.sciencedirect.com/science/article/pii/037702659390014X, 1993.

Hoke, J. and Anthes, R.: The Initialization of Numerical Models by a Dynamic Relaxation Technique, Mon. Weather Rev., 104, 15511556, 1976.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112 – 126, doi:10.1016/j.physd.2006.11.008, http://www.sciencedirect.com/science/article/pii/S0167278906004647, 2007.

Kalman, R. E. and Bucy, R. S.: New results in Linear Filtering and Prediction theory, Journal of Basic Engineering, 83, 95–108, doi:10.1115/1.3658902, 1961.

Kalnay, E. and Yang, S.-C.: Accelerating the spin-up of Ensemble Kalman Filtering, Quarterly Journal of the Royal Meteorological Society, 136, 1644–1651, doi:10.1002/qj.652, http://dx.doi.org/10.1002/qj.652, 2010.

Kalnay, E., Ki Park, S., Pu, Z., and Gao, J.: Application of the quasi-inverse method to data assimilation, Month. Weather Rev., 128, 864–875, 2000.

Killworth, P. D., Dieterich, C., Le Provost, C., Oschlies, A., and Willebrand, J.: Assimilation of altimetric data and mean sea surface height into an eddy-permitting model of the North Atlantic, Progress in Oceanography, 48, 313 – 335, doi:10.1016/S0079-6611(01)00009-X, http://www.sciencedirect.com/science/article/pii/S007966110100009X, 2001.

Krysta, M., E., B., Cosme, E., and Verron, J.: A consistent hybrid variational-smoothing data assimilation method: Application to a simple shallow-water model of the turbulent mid-latitude ocean, Month. Weath. Rev., 139, 3333–3347, 2011.

Lakshmivarahan, S. and Lewis, J.: Nudging Methods: A Critical Overview, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, edited by Park, S. K. and Liang, L., vol. II, pp. 27–58, Springer Verlag, Berlin, 2012.

Le Dimet, F. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations, Tellus, 38A, 97–110, 1986.

Leghtas, Z., Mirrahimi, M., and Rouchon, P.: Back and Forth Nudging for quantum state estimation by continuous weak measurement, in: Proceedings of American Control Conference, pp. 4334–4339, San Francisco, USA, 2011.

Lei, L., Stauffer, D., Haupt, S. E., and Young, G.: A hybrid nudging-ensemble Kalman filter approach to data assimilation. Part I: application in the Lorenz system, Tellus A, 64, 2012.

Leredde, Y., Devenon, J.-L., and Dekeyser, I.: Turbulent viscosity optimized by data assimilation, Annales Geophysicae, 17, 1463–1477, doi:10.1007/s00585-999-1463-9, http://www.ann-geophys.net/17/1463/1999/, 1999.

Levy, M. M., Klein, P., Trguier, A.-M., Iovino, D., Madec, G., Masson, S., and K. Takahashi, S.: Modifications of gyre circulation by sub-mesoscale physics, Ocean Modelling, 34, 1 – 15, doi:10.1016/j.ocemod.2010.04.001, http://www.sciencedirect.com/science/article/pii/S1463500310000582, 2010.

Lewis, J. K., Shulman, I., and Blumberg, A. F.: Assimilation of Doppler radar current data into numerical ocean models, Continental Shelf Research, 18, 541 – 559, doi:10.1016/S0278-4343(98)00006-5, http://www.sciencedirect.com/science/article/pii/S0278434398000065, 1998.

Li, H., Kanamitsu, M., and Hong, S.-Y.: California reanalysis downscaling at 10 km using an ocean-atmosphere coupled regional model system, Journal of Geophysical Research, doi:10.1029/2011JD017372, 2012.

Lingala, N., Sri Namachchivaya, N., Perkowski, N., and Yeong, H.: Optimal Nudging in Particle Filters, Procedia {IUTAM}, 6, 18 – 30, doi:10.1016/j.piutam.2013.01.002, http://www.sciencedirect.com/science/article/pii/S2210983813000035, 2013.

Lions, J. L.: Optimal Control of Systems Governed by Partial Differential Equations., Springer-Verlag, Berlin, Federal Replubic of Germany, first edn., 1971.

Luenberger, D.: An introduction to observers, IEEE Transactions on Automatic Control, 16, 596–602, 1971.

Luenberger, D. G.: Observers for Multivariable Systems., IEEE Transactions on Automatic Control, 11, 190–197, doi:10.1109/TAC.1966.1098323, 1966.

Luo, X. and Hoteit, I.: Ensemble Kalman filtering with residual nudging, Tellus A, 64, doi:10.3402/tellusa.v64i0.17130, http://www.tellusa.net/index.php/tellusa/article/view/17130, 2012.

Luo, X. and Hoteit, I.: Efficient particle filtering through residual nudging, Quarterly Journal of the Royal Meteorological Society, 140, n/a–n/a, doi:10.1002/qj.2152, http://dx.doi.org/10.1002/qj.2152, 2013.

Madec, G.: NEMO ocean engine, Note du Pole de modlisation, Institut Pierre-Simon Laplace (IPSL), Paris, France, 27 edn., 2008.

Marchesiello, P., McWilliams, J. C., and Shchepetkin, A.: Open boundary conditions for long-term integration of regional oceanic models, Ocean Modelling, 3, 1 – 20, doi:10.1016/S1463-5003(00)00013-5, http://www.sciencedirect.com/science/article/pii/S1463500300000135, 2001.

Mogensen, K., Balmaseda, M. A., Weaver, A. T., M., M., and Vidard, A.: NEMOVAR: A variational data assimilation system for the NEMO ocean model, ECMWF Newsletter, 120, 2009.

Molcard, A., Griffa, A., and Ozgokmen, T. M.: Lagrangian Data Assimilation in Multilayer Primitive Equation Ocean Models, J. Atmos. and Ocean Tech., 22, 70–83, 2004.

Morel, P., Lefevre, G., and Rabreau, G.: On initialization and non-synoptic data assimilation, Tellus A, 23, http://www.tellusa.net/index.php/tellusa/article/view/10496, 1971.

Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Mon. Weather Rev., 129, 1494–1207, 2001.

Primeau, F. W.: Multiple equilibria of a double-gyre ocean model with super-slip boundary conditions., J. Phys. Oceanography, 28, 2130–2147, 1998.

Pu, Z., Kalnay, E., Sela, J., and Szunyogh, I.: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method, Month. Weather Rev., 125, 2479–2503, 1997.

Ramdani, K., Tucsnak, M., and Weiss, G.: Recovering the initial state of an infinite-dimensional system using observer, Automatica, 2010.

Roullet, G. and Madec, G.: salt conservation, free surface, and varying levels : a new formulation for ocean general circulation models., J. Geophys. Res., 105, 23,927–23,942, 2000.

Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Mon. Wea. Rev., 132, 3019–3032, 2004.

Stauffer, D. and Bao, J.-W.: Optimal determination of nudging coefficients using the adjoint equations, Tellus A, 45, 358–369, 1993.

Stauffer, D. R. and Seaman, N. L.: Use of Four-Dimensional Data Assimilation in a Limited-Area Mesoscale Model. Part I: Experiments with Synoptic-Scale Data, Monthly Weather Review, 118, 1250–1277, doi:10.1175/1520-0493(1990)118⟨1250:UOFDDA⟩2.0.CO;2, http://dx.doi.org/10.1175/1520-0493(1990)118⟨1250:UOFDDA⟩2.0.CO;2, 1990.

Talagrand, O.: A study of the dynamics of four-dimensional data assimilation, Tellus, 33, 43–60, 1981.

Tenenhaus, M.: La régression PLS : Théorie et Pratique., éditions Technip, Paris, France, first edn., 1998.

Thompson, K. R., Wright, D. G., Lu, Y., and Demirov, E.: A simple method for reducing seasonal bias and drift in eddy resolving ocean models, Ocean Modelling, 13, 109 – 125, doi:10.1016/j.ocemod.2005.11.003, http://www.sciencedirect.com/science/article/pii/S1463500305000910, 2006.

Verron, J.: Nudging satellite altimeter data into quasi-geostrophic ocean models, Journal of Geophysical Research: Oceans, 97, 7479–7491, doi:10.1029/92JC00200, http://dx.doi.org/10.1029/92JC00200, 1992.

Vidard, P. A., Le Dimet, F.-X., and Piacentini, A.: Determination of optimal nudging coefficients, Tellus A, 55, 1–15, doi:10.1034/j.1600-0870.2003.201317.x, http://dx.doi.org/10.1034/j.1600-0870.2003.201317.x, 2003.

Wang, K., Debernard, J., Sperrevik, A. K., Isachsen, E., and Lavergne, T.: A combined optimal interpolation and nudging scheme to assimilate OSISAF sea-ice concentration into ROMS, Annals of Glaciology, 54, 8–12, 2013.

Weaver, A. and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation, Quarterly Journal of the Royal Meteorological Society, 127, 1815–1846, doi:10.1002/qj.49712757518, http://dx.doi.org/10.1002/qj.49712757518, 2001.

Weaver, A. T., Deltel, C., Machu, E., Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, Q. J. R. Meteorol. Soc, 131, 3605–3625, 2005.

Willians, P. D.: A Proposed Modification to the Robert Asselin Time Filter., Month. Weather Rev., 137, 2538–2546, 2009.

Zeitz, M.: The extended Luenberger observer for nonlinear systems, Systems and Control Letters, pp. 149–156, 1987.

Zou, X., Navon, I. M., and Le Dimet, F. X.: An Optimal Nudging Data Assimilation Scheme Using Parameter Estimation, Quarterly Journal of the Royal Meteorological Society, 118, 1163–1186, doi:10.1002/qj.49711850808, http://dx.doi.org/10.1002/qj.49711850808, 1992.