Nonlin. Processes Geophys. Discuss., 1, 1073–1131, 2014
www.nonlin-processes-geophys-discuss.net/1/1073/2014/
doi:10.5194/npgd-1-1073-2014

Nonlinear Processes in Geophysics
Discussions
Open Access

Data assimilation experiments using the diffusive back and forth nudging for the NEMO ocean model

G. A. Ruggiero[1], Y. Ourmières[2], E. Cosme[3], J. Blum[1], D. Auroux[1], and J. Verron[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), Marseille, France
[3]Université Joseph Fourier/LGGE, Grenoble, France
[4]CNRS/LGGE, Grenoble, France

Received: 14 May 2014 – Accepted: 19 June 2014 – Published: 16 July 2014

Correspondence to: G. A. Ruggiero (giovanni.ruggiero@unice.fr)

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

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.

1 Introduction

The well-known Nudging method is based on the second Newton axiom and consists in adding a forcing term in the right hand side of a given system in order to gently push the model toward a prescribed value. The first appearance of nudging in the geophysical literature was in 1974 (Anthes, 1974). In this work the authors proposed the use of nudging to mitigate initialization problems in atmospheric models. However, a similar algorithm had already been developed by Luenberger (1966). This algorithm has been called “Luenberger observer” or “asymptotic estimator”, since under linearity and observability hypothesis the estimator error converges to zero for time tending to

NPGD
1, 1073–1131, 2014

infinity. It is quite interesting to note that there is no mention of the Luenberger observer in the geophysical literature except in the recent work of Auroux and Blum (2005). 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 oceanographic 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 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. 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 huge state space (10^8 variables). This may require several iterations of the minimization algorithm, which in practice requires 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

NPGD
1, 1073–1131, 2014

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 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.

Recently, Auroux and Blum (2005) revisited the nudging method and proposed a new observer called Back and Forth Nudging (BFN). 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 the 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.

The BFN convergence was proved by Auroux and Blum (2005) for the linear case and full observations, 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).

NPGD
1, 1073–1131, 2014

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 but with all the middle latitude ocean mesoscale process features.

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 NEMO considers more vertical degrees of freedom, since the SW model has no vertical levels and the LQG was implemented with only 3 layers, as well as it 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

NPGD
1, 1073–1131, 2014

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 methodology and the gains configuration as well as of the 4Dvar and its configuration are described. Section 3 details the model set-up and the experimental design. Section 4 discusses the practical aspects of the backwards integration. The results are presented in Sect. 5, for which five aspects are analysed:

- sensitivity of the BFN to the assimilation window length (Sect. 5.1.1) ;
- sensitivity to the model diffusion coefficients (Sect. 5.1.2);
- effects of the number of iterations (Sect. 5.1.3);
- importance of the nudging gain structure (Sect. 5.2.1);
- impact of spatial and temporal observations distribution on the method performance (Sects. 5.2.2 and 5.2.3).

Finally, this section ends with a comparison between the BFN and 4Dvar for the case where observations are distributed so as to simulate an altimetry satellite track.

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

NPGD
1, 1073–1131, 2014

(or discretized partial differential equations), nudging consists in adding to them the forcing term $\mathbf{K}(\boldsymbol{x}_{\text{obs}} - \mathcal{H}(\boldsymbol{x}))$:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \mathcal{F}(\boldsymbol{x}) + \mathbf{K}(\boldsymbol{x}_{\text{obs}} - \mathcal{H}(\boldsymbol{x})) \quad (1)$$

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 $\boldsymbol{x}_{\text{obs}}(t)$ to the corresponding system state $\mathcal{H}(\boldsymbol{x})$, and $\mathbf{K}$ is the nudging gain matrix. In this algorithm the model appears as a weak constraint. The feedback term changes the dynamical equations and forces the state variables to fit the observations as well as possible.

In the linear case, i.e. when $\mathcal{F}$ and $\mathcal{H}$ may be written as matrices $\mathbf{F}$ and $\mathbf{H}$, 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 $(\mathbf{F}, \mathbf{H})$ holds, there is a class of possible values of $\mathbf{K}$ that guarantees the estimator convergence when $t \rightarrow \infty$ (Gelb et al., 1974). This should be one possible explanation why nudging usually works quite well and the converged state is not strongly affected by the choice of $\mathbf{K}$. However, when constructing $\mathbf{K}$ (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 BFN novelty with respect to conventional nudging methods 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. The iteration process makes possible the use of gains relatively weaker than those required by conventional nudging, thus enabling the conservation of physical constraints without affecting the estimator response.

NPGD
1, 1073–1131, 2014

Under the hypothesis of a linear model a variational interpretation is possible. In this case, if we choose $\mathbf{K} = k\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$, where $\mathbf{R}$ is the observation error covariance matrix, and k is a scalar, the solution of the estimation problem is a compromise between the minimization of the system's energy and the minimization of the distance between the data and the model (Auroux, 2009).

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).

Still with respect to the backward model diffusion, the loss/gain of energy in the forward integration should be compensated by an increase/decrease of energy in the correct spectral band, in the backward integration, without affecting numerical stability. This would be achieved by considering anti-diffusion (backward integration) followed by the application of a spectral filter in a similar way to Large Scale Eddy Simulation, or constructing an order N generalized diffusion operator, $\sum_{n=1}^{N}(-i)^n\nabla^n$ that could provide the expected response for the length scale of interest and suppress numerical noise. We are currently investigating both possibilities, but this remains beyond the scope of this paper.

Meanwhile in this work, the Diffusive Back and Forth Nudging-DBFN (Auroux et al., 2011) is used, for which 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. All the possible mitigatory solutions just described do not take into account water masses transformations resulting from diffusion processes. Indeed, unmixing water masses in the backward integration would be quite difficult and expensive.

NPGD
1, 1073–1131, 2014

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

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

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 Eq. (2) it gives:

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

where $k \in \mathbb{N}_{\geq 1}$ stands for iterations. Nudging applied to the backward system with the reversed diffusion sign gives:

$$\frac{\partial \tilde{\boldsymbol{x}}_k}{\partial t} = \mathcal{F}(\tilde{\boldsymbol{x}}_k) - \nu \Delta \tilde{\boldsymbol{x}}_k - \mathbf{K}'(\boldsymbol{x}_{\text{obs}} - \mathcal{H}(\tilde{\boldsymbol{x}}_k)) \tag{4}$$
$$\tilde{\boldsymbol{x}}_k(T) = \boldsymbol{x}_k(T), \quad T > t > 0.$$

The system composed by Eqs. (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:

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

NPGD
1, 1073–1131, 2014

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

According to Y. Brenier (personal communication, 2013), if $\mathbf{K}' = \mathbf{K}$ and the forward and backward limite trajectory are equal, i.e $\tilde{\boldsymbol{x}}_\infty = \boldsymbol{x}_\infty$, then $\boldsymbol{x}_\infty$ satisfies the model equations without the Nudging and diffusion:

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

as well as the Poisson equation:

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

which represents a smoothing process on the observations for which the degree of smoothness is given by the ratio $\frac{\nu}{\mathbf{K}}$ (Auroux et al., 2011).

Therefore, the DBFN provide estimations of the system which are smooth enough to avoid initialization problems related to the introduction of sparse and noisy observations into the system, at the same time they satisfy the model equation without diffusion.

In this study the matrix $\mathbf{K}$ is considered as a scalar matrix, i.e. a diagonal matrix with all its entries equal to a scalar (see Sect. 3.2, Table 1 for more details), and constructed using Regression Models (RM). Our choice for a diagonal gain is based on the encouraging results found by Auroux and Blum (2008) using the BFN with simplified ocean models. If $\mathbf{K}$ is interpreted as the Kalman gain (Gelb et al., 1974), it is diagonal when covariances between model variables are ignored and the covariance matrix of the observation errors is diagonal, the latter being a common assumption in DA applications (Pham, 2001; Brankart et al., 2010). In this case, only the observed part of the state space is directly controlled. Nevertheless, corrections of the non-observed variables are done by the model itself. In the case of a scalar matrix, it is further assumed that the error variance is the same for all observed state variables. Concerning regression models, the algorithm operates in two steps: first the observed variables are updated

NPGD
1, 1073–1131, 2014

and subsequently the other state variables are calculated using linear regression. The Partial Least Squares (PLS) regression (Tenenhaus, 1998) is used. The gain **K** is kept constant over the assimilation cycles. Our updating scheme can be seen as a rough approximation of the two steps update for EnKF presented by Anderson (2003).

Partial Least Squares regression (PLS)

The PLS was first introduced by Wold (1975) to address the problem of econometric path modeling, and was subsequently adopted for regression problems in chemometric and spectrometric modeling. In the method description, $\mathbf{X} \in \mathbb{R}^{n \times M}$ is considered as the observed or predictor variables and $\mathbf{Y} \in \mathbb{R}^{n \times N}$ 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 **X** and **Y**. Besides, **X** and **Y** are centered and have the same units. The PLS regression features two steps: a dimension reduction step in which the predictors from matrix **X** 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 **X**:

$$\mathbf{T} = \mathbf{XW}, \tag{8}$$

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

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

$$\boldsymbol{w}_i = \mathrm{argmax}_{\boldsymbol{w}} \{\mathrm{cov}(\mathbf{X}\boldsymbol{w}, \mathbf{Y})^2\} \tag{9}$$

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

NPGD
1, 1073–1131, 2014

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

$$\hat{\mathbf{B}}^{\mathrm{PLS}} = \mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \quad (10)$$

An immediate consequence of Eq. (10) is that when $\mathbf{W} = \mathbf{I}$ 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.

2.2 Four dimensional variational method – 4DVar

The objective of the variational methods is to 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 $(t_i)_{1\leq i\leq N}$. 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:

$$J(\boldsymbol{x}_0) = \frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i[\mathcal{M}_{0,i}(\boldsymbol{x}_0)] - \boldsymbol{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i[\mathcal{M}_{0,i}(\boldsymbol{x}_0)] - \boldsymbol{y}_i) \quad (11)$$

where $\mathbf{B}$ is the background error covariance matrix and $\mathcal{M}_{0,i}$ represents the model integration from time t_0 to time t_i. $\mathbf{R}_i$,$\mathcal{H}_i$ 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:

$$\nabla J(\boldsymbol{x}^a(t_0)) = 0. \quad (12)$$

NPGD
1, 1073–1131, 2014

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).

If $\mathcal{H}$ or $\mathcal{M}$ are non-linear, the solution of the problem is not unique, i.e. the functional Eq. (11) may have multiple local minima, and the minimization procedure may not stop at the global minimum. To overcome this problem, Courtier et al. (1994) proposed to solve a sequence of quadratic problems, expecting this sequence would converge to the solution of the problem given by Eqs. (11) and (12). This algorithm is called the incremental 4Dvar. 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:

$$\mathcal{M}_{0,i}(\boldsymbol{x}^b + \delta \boldsymbol{x}_0) \approx \mathcal{M}_{0,i}(\boldsymbol{x}^b) + \mathbf{M}_{0,i}\delta \boldsymbol{x}_0 \quad \forall i \tag{13}$$

$$\mathcal{H}_i(\boldsymbol{x}^b + \delta \boldsymbol{x}_0) \approx \mathcal{H}_i(\boldsymbol{x}^b) + \mathbf{H}_i\delta \boldsymbol{x}_0 \quad \forall i \tag{14}$$

and the new cost function is given by:

$$J(\delta \boldsymbol{x}_0) = \frac{1}{2}\delta \boldsymbol{x}_0^{\mathrm{T}}\mathbf{B}^{-1}\delta \boldsymbol{x}_0 + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\mathbf{M}_{0,i}\delta \boldsymbol{x}_0 - \boldsymbol{d}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_{0,i}\delta \boldsymbol{x}_0 - \boldsymbol{d}_i) \tag{15}$$

where $\boldsymbol{d}_i = \boldsymbol{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\boldsymbol{x}_b))$ is called the innovation vector. It is possible that after some iterations of the minimizer the increments become too large and a new linearization of $\mathcal{H}$ and $\mathcal{M}$ should be done. 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. See below a simplified description of this algorithm:

- **Initialization**: $\boldsymbol{x}_0^0 = \boldsymbol{x}^b$
- **While** $k \leq k_{\max}$ or $\|\delta \boldsymbol{x}_0^{a,k}\| > \epsilon$ (**Outer Loop**)

  Do

NPGD
1, 1073–1131, 2014

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

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

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

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

The **B** matrix used in this study was built following Weaver et al. (2005). In this formulation the matrix is decomposed as $\mathbf{B} = \mathbf{G}\boldsymbol{\Lambda}^{\mathrm{T}}\mathbf{C}\boldsymbol{\Lambda}\mathbf{G}^{\mathrm{T}}$, where **G** is a multivariate balance operator, **Λ** is a diagonal matrix of error variance, for which the climatological variances are the entries, and **C** is a univariate correlation matrix modeled using the generalized diffusion equation. The balance operator is meant to propagate information from the observed variable to the non-observed variables. It is composed by a set of linear and non-linear relationships between the state variables such as the geostrophic balance and some temperature and salinity constraints, for example. The matrix **R** is diagonal.

3 Ocean model and experimental set-up

The ocean model used in this study is the ocean component of NEMO (Nucleus for European Modeling 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).

NPGD
1, 1073–1131, 2014

The model solves six prognostic equations, namely the momentum balance, the hydrostatic equilibrium, the incompressibility equation, the heat and salt conservation equations and a nonlinear equation of state which 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.

Since Sea Surface Height (SSH) is one of the most informative ocean observation and is the variable to be assimilated in our experiments, it is particularly interesting to look at the model's pressure gradient formulation. As the free surface formulation is used, the pressure at a given point (x,y,Z,t) is given by $p(x,y,Z,t) = \int_0^Z g\rho(x,y,z,t)\mathrm{d}z + \rho_0 g\eta(x,y,t)$, where η is the free surface and describes the perturbation of pressure in relation to geopotential height $Z = 0$. Thus, p is the sum of the hydrostatic pressure and the surface pressure (depth independent).

While the hydrostatic part depends directly on the density field distribution, the free surface evolves according to:

$$\frac{\partial \eta}{\partial t} = -\nabla_\mathrm{h}(D\overline{U}_\mathrm{h}) + E - P$$

$$\overline{U}_\mathrm{h} = \frac{1}{D}\int_{-D}^{0} u_\mathrm{h}\mathrm{d}z, \tag{16}$$

where D is the water depth, $E - P$ is the evaporation-precipitation balance, u_h is the horizontal components of the velocity field and ∇_h is the divergence operator restricted

NPGD
1, 1073–1131, 2014

to the horizontal plane. In our experiments, the evolution of η depends only on the divergence of the vertically integrated velocities since the $E - P$ balance is set to zero. Equation (16) shows that resolving the vertical structure of the velocity field given observations of η is an underdetermined problem.

Analytical and numerical studies as well as direct observations have shown that most of the ocean variability corresponds to motions associated to the barotropic and the first baroclinic modes (Wunsch, 1997). This fact has been usually used to constrain DA solutions to the space spanned by these modes. This may be done by using an iterative procedure in which water masses are re-distributed on the vertical (Cooper and Haines, 1996), by ensemble methods which statistically capture these modes (Gavart and De Mey, 1997) or by explicitly restricting the minimization of cost functions similar to that given by Eq. (11) to the space spanned by theses modes.

Therefore, one question to be addressed in this study is: if nudging is applied to the free surface height, which corresponds to directly control the barotropic mode, are the baroclinic fields also satisfactorily corrected?

3.1 Ocean model configuration

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 25 km resolution and a vertical resolution ranging from 100 m near the upper surface to 500 m 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 $\tau =$

NPGD
1, 1073–1131, 2014

$(\tau_0 \cos\left(\frac{2\pi(y-y_0)}{L}\right), 0)$, where y is the latitude geographic coordinate with $y_0 = 24^\circ$ and $y_0 \le y \le 44^\circ$, $L = 20^\circ$ and $\tau_0 = -0.1\,\mathrm{N\,m^{-2}}$. 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: $\nu_h^{u,v} = -8 \times 10^{10}\,\mathrm{m^4\,s^{-1}}$ and $\nu_v^{u,v} = 1.2 \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$ for the momentum equations and $\nu_h^{t,s} = -4 \times 10^{11}\,\mathrm{m^4\,s^{-1}}$ and $\nu_v^{t,s} = 1.2 \times 10^{-5}\,\mathrm{m^2\,s^{-1}}$ 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 35 psu and a temperature field created to provide a stratification which has a first baroclinic deformation radius of 44.7 km. Velocity and pressure 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). Concerning this article, 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 and then two additional years were finally completed to be used as the truth, from which the observations were extracted.

3.2 Data assimilation configuration

This study addresses five main aspects of the DBFN: (i) sensitivity to the length of the assimilation window, (ii) sensitivity to the model diffusion coefficients, (iii) effects of the number of iterations, (iv) importance of the nudging gain structure and (v) impact of spatial and temporal observations distribution on the performance.

A first set of experiments, summarized in Table 1 and presented in Sect. 5.1, are designed to cover the first three main aspects. The data used are daily SSH fields available at every grid point and perturbed with a Gaussian white noise with a signal-to-noise ratio of 20 %. The SSH assimilation at every grid point would be similar to assimilate gridded products such as those produced by AVISO (www.aviso.oceanobs.com)

NPGD
1, 1073–1131, 2014

as done by Zavala-Garay et al. (2012). The difference in this case is that the observation errors may be correlated and in our case they are independent. The nudging gain is a scalar chosen to be strong enough to control the model errors but without being the dominant term of the equations.

More specificaly, in order to assess the BFN performance with respect to the length of the assimilation window (Sect. 5.1.1) five assimilation windows were tested: 2, 5, 10, 20 and 30 days. For each configuration, two different values for the diffusion coefficients were considered. The default value and a reduced value chosen on basis of Sect. 4 results. The results produced with the reduced diffusion are presented in Sect. 5.1.2. The effects of the number of iterations are analysed in Sect. 5.1.3 by comparing the results produced using the convergence criterion to those limiting the number of iterations to two.

In Sect. 5.2 the last two main aspects are analysed. In Sect. 5.2.1 the experiment ssh_10d_rd, which assimilates the SSH observations employing a 10 days assimilation window and a reduced diffusion coefficient, is compared to an experiment employing the same configuration but with the nudging gain **K** based on the PLS regression mode. In this case small increments produced by the regression model are propagated by the non-linear model forward and backward. Afterwards in Sect. 5.2.2, the impacts of the **K** structure is assessed, when observations are available every four days and subsequently, in Sect. 5.2.3 it is considered that every four days a sampling similar to the Jason-1 satellite is available. In this case the results produced by the DBFN are compared to the ones produced by the 4Dvar.

When the gain **K** is diagonal and with daily observations, a linear interpolation of observations is used to make observations available at every time step. In the case where observations are available every four days, a linear weighting function that decreases to zero in two days is used to weight the gain **K** while using the last available observation. When the PLS is considered no weighting function is used.

As we only work with simulated data (extracted from a reference trajectory), the methods' performance is assessed by analyzing the global relative error calculated as

NPGD
1, 1073–1131, 2014

$\frac{\|x-x^{true}\|}{\|x^{true}\|}$. For the 3-dimensional fields the relative error for each layer is also presented. With this approach, the performance of each experiment can be analysed with respect to the vertical structure.

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:

$$\text{Rerror} = \frac{\|\boldsymbol{x}(0) - \tilde{\boldsymbol{x}}(0)\|}{< \|\boldsymbol{x}(t + \Delta t) - \boldsymbol{x}(t)\| >} \tag{17}$$

where Δt = 1 day and the brackets represent the empirical mean.

Figure 1 shows the global error, Rerror, 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

NPGD
1, 1073–1131, 2014

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.

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

Numerical errors were assessed by changing the model time step from 900 s to 90 s. 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

NPGD
1, 1073–1131, 2014

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 190 km, 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 spectra obtained from the backward integration is observed at 126, 314 and 627 km 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 314 km 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 126 to 25 km. 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.

The spectral differences may be due to two things: first due to the modified energy flux between the different scales, which is verified by a change of the spectrum slope especially at high wave number, and second due to the smoothing of the large scale gradients.

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.

NPGD
1, 1073–1131, 2014

It has been observed that diapycnal mixing (i.e. mixing between two isopycnal surfaces) are several order of magnitude smaller than mixing within an isopycnal layer. NEMO is a Z coordinate model and the diffusion operators are split into vertical and horizontal components. This means that the problem of unrealistic diapycnal mixing, due to horizontal Z layers intersecting isopycnal plans, is even more pronounced due to the iterations. Therefore, the use of diffusion operators acting along isopycnals can be of great interest along with using the DBFN.

In the next section data assimilation ability to control the errors induced by diffusion is assessed as well as the consequences of these errors on the assimilation system configuration.

5 Data assimilation results

5.1 Experiments with scalar nudging coefficients

5.1.1 Sensitivity to the DA window length

Following the results presented in the previous sections, the results concerning the sensitivity test of the DBFN to the choice of the DA window are presented, but this time including the nudging term. Thus, the considered primary aspect is whether the use of a diagonal nudging gain is sufficient to control the backward model in the presence of anti-diffusion. In this case, control was achieved using an unrealistic observation network combined with strong nudging and for short time periods ($\leq$ 20 days). Indeed, the noise induced by anti-diffusion can potentially damage the mass field since it induces large rates of spurious dyapicnal mixing.

Figure 5 shows the zonal velocity and the temperature errors for the DBFN experiments assimilating the SSH. The experiments were integrated for 160 days, and five different DA windows (2, 5, 10, 20 and 30 days) were considered. All experiments used the same nudging gain and the same convergence criterion.

NPGD
1, 1073–1131, 2014

For all experiments the DBFN reduces the initial error by more than 60 % for the dynamical variables and by more than 30 % for the temperature. Long assimilation windows produce better results for the beginning of the experiments (first 30 days), which is a consequence of the asymptotic behavior of the algorithm, but using a short assimilation window is preferable at long term. The latter may be verified in Table 2, which presents the mean initial and final condition errors calculated for the entire experiments.

Moreover, for all cases the temperature error is not asymptotically stable, and starts to grow after 100 days. The experiment divergence is especially large for long assimilation windows due to the cumulative effect of the diffusion-induced errors. This is further amplified in our experiments since our model does not consider any heat source and therefore the balance between the downward heat diffusion and the upward cold water advection, which is responsible to the maintenance of the thermocline, is disturbed. Also, the results prove the difficulty of controlling the diffusion-induced errors by only assimilating SSH observations.

Futhermore, as observed by Reynolds and Palmer (1998) examining the accuracy of a backward tangent linear quasi-geostrophic model, the backward model spectrum whithout nudging may have lower error growth rates than the forward model when the diffusion sign is reversed, as in our case, explaining the good performance of the DBFN at the beginning of the experiments for all state variables. To explain the long term divergence, we call attention to the fact that in the absence of observations the DBFN produces fields that satisfy the Laplace equation, i.e it produces homogeneous fields which is obviously not desired.

In real applications, the SSH is assimilated together with vertical temperature profiles and/or sea surface temperature. Including the vertical profiles is enough to control the temperature drift observed in this section. Another possibility which is successfully applied in Sect. 5.2.3 is to estimate the temperature from the SSH using regression models.

NPGD
1, 1073–1131, 2014

5.1.2 Sensitivity to the diffusion coefficient

In the last section it was shown that even with nudging the backward error is quite large. To study the DBFN sensitivity to the model diffusion coefficient, the experiments from Sect. 5.1.1 are repeated but use the diffusion coefficients that gave the best results showed in Sect. 4.

By comparing Tables 2 and 3, it is evident that the use of a reduced diffusion improves the state estimation for all variables and window sizes. For instance, Figs. 5 and 6 reveal that after 160 days of the experiments and considering an assimilation window of 10 days, the velocity error is almost 45 % smaller when considering a reduced diffusion.

Again, it seems to be preferable to consider short assimilation windows. The velocity errors are stable and smaller for 5 and 10 days windows. For the temperature, the shorter the assimilation window, the smaller the errors. They also initially decrease but start to grow after 100 days, excepted for the 2 days window. This confirmes the importance of the cumulative effect of the diffusion-induced errors.

Reducing the diffusion coefficient modifies the model singular spectrum. For this reason we compare in the Table 4 the forecast error growth rate for the experiments employing the reference and the reduced diffusion coefficients. For almost all experiments the growth rate is smaller when using the reference diffusion coefficient.

This finding supports the discussion of the last section about the backward model accuracy. If on the one hand, a reduced diffusion improves the backward model accuracy, on the other hand the error growth rate, both in forward and backward, may increase, leading to large forecast errors. This happens because all singular values of the model increase as dissipation is decreased (Reynolds and Palmer, 1998).

A simple solution, but not tested here, would be to use distinct diffusion coefficients in the forward and backward integrations. Furthermore, the convex character of the error curves is pronounced for long assimilation windows and reduced diffusion experiments, revealing the importance of the unstable portion of the spectrum for both the forward

NPGD
1, 1073–1131, 2014

and the backward integrations. This pattern was also observed by examining the error evolution as iterations are performed (not shown).

5.1.3 Effects of the number of iterations and the control of the ocean vertical structure

Two new sets of experiments have been created using a reduced diffusion and assimilating the SSH. For the first, the number of iterations is limited to two while for the second the convergence criterion is more restrictive ($\epsilon = 0.001$) and iterations are limited to 50. Starting from the latter, Fig. 7 shows that if only one assimilation cycle is considered, more iterations results in more accurate initial condition with respect to the dynamical variables. For the temperature, the estimation starts to diverge after 8, 10 and 20 iterations for the 30, 20 and 10 days assimilation window respectively. In addition, the 10 days window has the smaller velocity error after convergence. However, as already noted in Sect. 5.1.1 the system diverges when several assimilation cycles are considered, regardless the assimilation window used.

Considering the convergence criterion $\epsilon = 0.005$, the DBFN converges quite fast (2–3 iterations) for all experiments. Significant differences are observed for the first three assimilation cycles for which much more iterations (7–20) are required. As shown in Fig. 7, the number of iterations for the first assimilation cycle depends on the length of the assimilation window. A long assimilation window leads to a faster convergence in terms of iterations. Nonetheless considering the computational cost, the short assimilation window would be preferable since the gain from reducing the number of iterations is not large enough to offset the computational cost of using longer assimilation windows.

When convergence is reached, the errors are different and vary with the variables, e.g. for the SSH the 2 days window provides the best result while for the zonal velocity and temperature the 20 and 10 days windows performed better, respectively. The apparent lack of pattern of the observed errors for different data assimilation windows

NPGD
1, 1073–1131, 2014

shown in the Fig. 7 must be related to the precision with which we estimate the mass field, and its influence in the estimation of the tridimensional velocity field.

In the case of two iterations (Fig. 8), no tendency is observed for the temperature errors for windows shorter than 20 days. However, velocity errors are larger than the errors obtained using the convergence criterion. When analyzing only the first assimilation window, the best results were obtained with the 30 days experiment, although it is the 10 days window followed by the 5 and 2 days one that produced the best mean initial conditions, see Table 3. The best results observed for the longer assimilation window is a consequence of the asymptotic character of the Nudging method. Therefore, the longer the assimilation window, the smaller the error. The same is not observed when several assimilation cycles are considered due to the diffusive aspect of the DBFN. Again, we recall that the Eq. (7) shows that in the absence of observations the DBFN solution at convergence will be a homogenous field. This explains the observed problems in estimating temperature especially for long assimilation windows and restrictive convergence criterions.

These results suggest that there is a trade-off between considering long assimilation windows that permit the use of the non-linearities and the correction of the errors projecting onto the stable subspace, and the errors due to diffusion. The presented experiments clearly indicate that for the observation network we have been considering the 10 days window fulfills this criterion. Therefore, in the following sections only the 10 days window is considered.

The vertical error structure is shown in Figs. 9 and 10 for the first and last assimilation cycle. The errors are reduced especially at the upper layers including the thermocline. This suggests that the model reproduced the appropriate dynamics since the thermocline dynamics is mainly controlled by the first baroclinic mode and recalling that assimilating the SSH is a direct control of the barotropic mode. The increase of temperature errors after 100 days is observed at depths where stratification is stronger, confirming the influence of diffusion in these errors structures. For the velocity, there is an increase of the error at depths larger than 1500 m. This may happen because

NPGD
1, 1073–1131, 2014

at this depth velocities may be preferably baroclinic and our nudging scheme corrects the barotropic mode, and/or because the excessive smoothing of the horizontal density gradients alters the velocity fields by changing the thermal wind balance and therefore the baroclinic velocities.

5.1.4 Conclusions

In this section we have seen that a relatively short assimilation window ($\leq$ 10 days) along with a reduced diffusion coefficient has to be preferably used with the DBFN, confirming the results presented in Sect. 4. Concerning the velocities, Figs. 2 and 10 show that nudging the SSH reduces the errors exactly where diffusion errors are larger, i.e. at the upper ocean. Controlling the deep ocean by assimilating only SSH is quite a difficult task. At this point we cannot make conclusions about the nature of the remaining errors, i.e. if they are caused by diffusion or if the SSH is not a good predictor for the deep ocean. This discussion continues in the Sect. 5.2.1 where the results produced by the regression models are discussed.

It can also be concluded that in the absence of a dense observational network, (e.g. sampling the ocean 3-dimensional structure), more complex gains, **K**, which correct the non-observed variables are needed. This is mostly due to the diffusive character of the algorithm. In order to consider this aspect, the use of gains based on regression models are analyzed in the following Sect. 5.2.

5.2 The Hybrid DBFN

In this section the importance of the nudging gain structure under different observation network is analysed. This is done by comparing experiments using a diagonal **K** and a **K** constructed using the PLS regression model. The experiments are summarized in Table 5.

NPGD
1, 1073–1131, 2014

5.2.1 Daily gridded SSH observations

In this section the experiments ssh_10d_rd and ssh_10d_rd_it2, which assimilates daily SSH fields with reduced diffusion coefficients and the experiments ssh_10d_rd_uv and ssh_10d_rd_it2_uv are compared. Their configuration are similar but with **K** constructed using the PLS regression model to correct velocity with SSH increments. This choice is based on results not presented in this article that show that good estimates of the velocities implies good estimates of the temperature field, and on the fact that the SSH is better correlated with the velocity field than with temperature.

The use of the regression model improves the estimation of all model variables when only two iterations were considered (Fig. 11). When the convergence criterion was used, the velocity errors are smaller for the first 4 DA cycles, after which the pure DBFN errors become slightly smaller than those of the experiment using the regression. Temperature estimation is also improved. This may be related to the amelioration of the upper ocean velocities.

The vertical structure of the velocity error for the day 1 and for the day 130 is analysed (Fig. 12) in terms of the Root Mean Squared Error and by decomposing the vertical error into empirical orthogonal functions (EOF). This decomposition shows what kind of errors remains after the assimilation step. For all experiments the first mode accounts for more than 90% of the error variance and represents an error which has the same sign over all depths. More specifically, the velocities over most of the domain are underestimated, showing again the importance of the model diffusion in extracting energy from the system. Independently of the matrix **K** used, after two iterations the surface velocities have higher errors than the deep ocean velocities. As iterations are performed the surface errors decrease faster than the deep errors. The gain **K** constructed using the PLS is clearly more effective in reducing the surface errors than the diagonal **K**. However due to the increasing vertical variance of the PLS estimator, the deep ocean structures are degraded as iterations are performed. This is especially

NPGD
1, 1073–1131, 2014

important in the regions where the dynamic is more intense, which indirectly means where the increments are bigger.

The similarity of the EOF modes between the experiments using a diagonal gain and that using the PLS indicates that both methods correct the model in the same way. The PLS advantage is its efficiency in improving the upper ocean estimation with respect to the DBFN, which is specially important when less iterations are considered. However, the PLS is not so efficient as the DBFN in reducing the deep ocean errors. This proves the DBFN skill in estimating the dynamical variables when gridded observations of SSH are available. Nevertheless, the improvement of the deep ocean velocities would require direct observations of this region, since the SSH does not appears to be a good predictor of this region.

5.2.2 Temporal data sparsity: gridded SSH obs. available every 4 days

In this section the experiments presented above have been reproduced but with observations available every 4 days (see experiments ssh4_10d_rd_uv, ssh4_10d_rd, ssh4_10d_rd_it2_uv and ssh4_10d_rd_it2 in Table 5). Temporal sparsity is in general a great challenge for nudging methods since they usually correct the observed variables but nothing is done with respect to the non-observed part of the state vector.

Figure 13 shows the results with and without the use of the regression model. In this case the errors of the experiments using the PLS regression are 40 % smaller for the velocities, 60 % for the SSH and 15 % for the temperature. Once again, using only 2 iterations is beneficial to keep the temperature errors stable. However, as shown in Fig. 12, iterations are responsible for the correction of the upper ocean velocities, resulting in better initial conditions and a more stable predictive step.

Therefore, we have seen that the use of **K**s accounting for corrections of the non-observed part of the state vector is mandatory in situations where the observational network is poor, as it is the case in real ocean applications.

The results presented in the previous sections are used in the following part to configure an experiment that uses a more realistic observation network.

NPGD
1, 1073–1131, 2014

5.2.3 Intercomparisons

In this section we assume that every four days an observation network simulating Jason-1 satellite sample is available. In addition, to validate the results produced by the DBFN, a comparison with the 4Dvar method is presented. Also, the temperature is included in the regression model. This makes the comparison with the 4Dvar more equitable, since the non-observed variables in 4Dvar are estimated by considering the multivariate balance operator **G** (see Sect. 2.2).

First the minimization performance of the 4Dvar implementation is analysed. Figure 14 shows the reduction of the cost function gradient 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 + PLS to converge, i.e., after which the errors cease to decrease. Moreover, the 4Dvar numerical cost is more than 3 times the DBFN + PLS cost since one execution of the adjoint model costs four times the cost of the direct model in terms of CPU time.

Figure 15 shows the relative error for the control experiment (without assimilation), an experiment using the direct nudging with PLS regression (ONDG), the DBFN + PLS and a 4Dvar. The DBFN + PLS experiment error is stable throughout the experiment 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 + PLS produced the best results for all variables, except for the zonal velocity, for which the 4Dvar has slightly smaller errors. The ONDG also showed good performance, but with mean errors larger than the DBFN + PLS and 4Dvar mean errors. Also, it has initialization problems that can be identified in the Fig. 15 as large oscillations after each assimilation step.

In terms of vertical error (Fig. 16), the DBFN + PLS and the ONDG performed better for the upper ocean than the 4Dvar. Clearly, the PLS also corrects the deep ocean velocity, but less accurately than the 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

NPGD
1, 1073–1131, 2014

4Dvar, 96 and 93 % for DBFN + PLS and ONDG, respectively. Although the first mode is the barotropic one for all methods, it is important to note that the 4Dvar barotropic mode 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 PLS 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 velocities are overestimated due to spurious covariances between the SSH and the deep ocean velocities. Therefore, the first mode variability produced by the ONDG and the DBFN + PLS confirms the role of iterations in improving the baroclinic fields.

The way both methods correct the model depends on the **B** matrix in the 4Dvar algorithm and on the latent structures in the DBFN + PLS. It means that results may be different if their calculation is changed. The main aspect of the results is that with a method which is easier to implement and cheaper to execute we can produce results which are at least equivalent 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 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 is relatively weak in relation to the dominant term of the equations.

6 Conclusions and perspectives

This study used the NEMO general circulation model and the double gyre configuration to investigate the Diffusive Back and Forth Nudging performance under different configurations of the data assimilation window, nudging gains and observations network, as well as to compare the DBFN with the 4Dvar.

NPGD
1, 1073–1131, 2014

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 can be used for ocean data assimilation despite the low accuracy of the backward integration. Improving the backward integration would further improve the DBFN performance and make possible the use of longer assimilation windows.

The use of scalar gains in the DBFN requires high spatial and temporal availability of data, otherwise, the method does not bring significant improvements due to the model inefficiency in spreading the observation information and hence the impossibility of controlling the diffusion errors. In the case of sparse data, complex functions are necessary to propagate the information from the data to the non-observed variables and to regions of the domain that are not observed. In this work, these functions were constructed using the PLS latent structures resulting from the process of maximizing the covariance between the observed variables and the variables to be estimated. In our implementation the PLS model does not vary temporally which makes the method less computationally demanding. Notably, the iterations are responsible for giving a dynamic character to the solution since the increments are propagated using the nonlinear model.

Our results show that the DBFN can produce results comparable with 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 twin experiment framework is favorable to the perfect model assumption used by our 4Dvar. However, the 4Dvar had difficulties to fit the SSH observations available in one assimilation window. This is observed especially at the beginning of the

NPGD
1, 1073–1131, 2014

experiment, when the background state is far from the observations. This can be an indicator of conditioning problems. The DBFN, however, does not suffer from this kind of problem; the errors on the dynamical variables reach their asymptotic values after the first assimilation cycle.

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 **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 from the MEOM team, University of Grenoble, for his help with the set-up of the 4Dvar algorithm.

References

Abarbanel, H. D. I., Kostuk, M., and Whartenby, W.: Data assimilation with regularized nonlinear instabilities, Q. J. Roy. Meteorol. Soc., 136, 769–783, doi:10.1002/qj.600, 2010. 1076

Anderson, J. L.: A local least squares framework for ensemble filtering, Mon. Weather Rev., 131, 634–642, 2003. 1083

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, 1974. 1074

Auroux, D.: The back and forth nudging algorithm applied to a shallow water model, comparison and hybridization with the 4D-VAR, Int. J. Numer. Meth. Fl., 61, 911–929, 2009. 1076, 1077, 1080

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. 1075, 1076

Auroux, D. and Blum, J.: A nudging-based data assimilation method: the Back and Forth Nudging (BFN) algorithm, Nonlin. Processes Geophys., 15, 305–319, doi:10.5194/npg-15-305-2008, 2008. 1076, 1077, 1082

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. 1076

NPGD

1, 1073–1131, 2014

Auroux, D., Blum, J., and Nodet, M.: Diffusive Back and Forth Nudging algorithm for data assimilation, C. R. Acad. Sci. Paris Ser. I, 349, 849–854, 2011. 1080, 1082

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. 1076

Ballabrera-Poy, J., Kalnay, E., and Yang, S.-C.: Data assimilation in a system with two scales – combining two initialization techniques, Tellus A, 61, 539–549, doi:10.1111/j.1600-0870.2009.00400.x, 2009. 1076

Bergemann, K. and Reich, S.: A mollified ensemble Kalman filter, Q. J. Roy. Meteorol. Soc., 136, 1636–1643, doi:10.1002/qj.672, 2010. 1076

Blayo, E., Verron, J., and Molines, J.-M.: Assimilation of topex/poseidon altimeter data into a circulation model of the north atlantic, J. Geophys. Res.-Oceans, 99, 24691–24705, 1994. 1075

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, 385–442, Elsevier, Oxford, UK, 2008. 1075

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. 1089

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, 0027–0644, doi:10.1175/2010MWR3310.1, 2010. 1082

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. Oceanogr., 31, 1260–1286, 2001. 1088

Chassignet, E. P. and Gent, P. R.: The influence of boundary conditions on midlatitude jet separation in ocean numerical models, J. Phys. Oceanogr., 21, 1290–1299, 1991. 1088, 1089

Chen, X., Liu, C., O'Driscoll, K., Mayer, B., Su, J., and Pohlmann, T.: On the nudging terms at open boundaries in regional ocean models, Ocean Model., 66, 14–25, doi:10.1016/j.ocemod.2013.02.006, 2013. 1075

NPGD
1, 1073–1131, 2014

Clifford, M., Horton, C., Schmitz, J., and Kantha, L. H.: An oceanographic nowcast/forecast system for the Red Sea, J. Geophys. Res.-Oceans, 102, 25101–25122, doi:10.1029/97JC01919, 1997. 1076

Cooper, M. and Haines, K.: Altimetric assimilation with water property conservation, J. Geophys. Res.-Oceans, 101, 1059–1077, doi:10.1029/95JC02902, 1996. 1088

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 Model., 33, 87–100, 2010. 1088

Courtier, P., Thepaut, J. N., and Hollingsworth, A.: A strategy for operational implementation of 4d-var, using an incremental approach., Q. J. Roy. Meteorol. Soc., 123, 1367–1387, 1994. 1075, 1085

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, Limassol, Cyprus, 1–5,2010. 1076

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res.-Oceans, 99, 10143–10162, doi:10.1029/94JC00572, 1994. 1075

Gavart, M. and De Mey, P.: Isopycnal EOFs in the Azores current region: a statistical tool for dynamical analysis and data assimilation, J. Phys. Oceanogr., 27, 2146–2157, 1997. 1088

Gelb, A., Kasper, J., Nash, R. A., Price, C. F., and Sutherland, A. A.: Applied Optimal Estimation, arthur gelb Edn., M.I.T. Press, Reading, Massachusetts, 1974. 1079, 1082

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, Dynam. Atmos. Oceans, 17, 89–133, doi:10.1016/0377-0265(93)90014-X, 1993. 1075

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, doi:10.1016/j.physd.2006.11.008, 2007. 1075

Kalnay, E., Ki Park, S., Pu, Z., and Gao, J.: Application of the quasi-inverse method to data assimilation, Mon. Weather Rev., 128, 864–875, 2000. 1080

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, Prog. Oceanogr., 48, 313–335, doi:10.1016/S0079-6611(01)00009-X, 2001. 1075

NPGD
1, 1073–1131, 2014

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, Mon. Weather Rev., 139, 3333–3347, 2011. 1088

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, Springer Verlag, 27–58, 2012. 1075

Le Dimet, F. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations, Tellus A, 38, 97–110, 1986. 1075, 1085

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, San Francisco, USA, 4334–4339, 2011. 1076

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, 18485, doi:10.3402/tellusa.v64i0.18485, 2012. 1076

Leredde, Y., Devenon, J.-L., and Dekeyser, I.: Turbulent viscosity optimized by data assimilation, Ann. Geophys.-Italy, 17, 1463–1477, doi:10.1007/s00585-999-1463-9, 1999. 1080

Levy, M. M., Klein, P., Tréguier, A.-M., Iovino, D., Madec, G., Masson, S., and Takahashi, K. S.: Modifications of gyre circulation by sub-mesoscale physics, Ocean Model., 34, 1–15, doi:10.1016/j.ocemod.2010.04.001, 2010. 1088

Lewis, J. K., Shulman, I., and Blumberg, A. F.: Assimilation of Doppler radar current data into numerical ocean models, Cont. Shelf. Res., 18, 541–559, doi:10.1016/S0278-4343(98)00006-5, 1998. 1075

Li, H., Kanamitsu, M., and Hong, S.-Y.: California reanalysis downscaling at 10 km using an ocean–atmosphere coupled regional model system, J. Geophys. Res., 117, D12118, doi:10.1029/2011JD017372, 2012. 1075

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, 2013. 1076

Lions, J. L.: Optimal Control of Systems Governed by Partial Differential Equations, 1st Edn., Springer-Verlag, Berlin, Germany, 1971. 1085

Luenberger, D. G.: Observers for multivariable systems, IEEE T. Automat. Contr., 11, 190–197, doi:10.1109/TAC.1966.1098323, 1966. 1074, 1079

Luo, X. and Hoteit, I.: Ensemble Kalman filtering with residual nudging, Tellus A, 64, 17130, doi:10.3402/tellusa.v64i0.17130, 2012. 1076

NPGD
1, 1073–1131, 2014

Luo, X. and Hoteit, I.: Efficient particle filtering through residual nudging, Q. J. Roy. Meteorol. Soc., 140, 557–572, doi:10.1002/qj.2152, 2013. 1076
Madec, G.: NEMO ocean engine, Note du Pole de modélisation, 27th Edn., Institut Pierre-Simon Laplace (IPSL), Paris, France, 2008. 1077
Marchesiello, P., McWilliams, J. C., and Shchepetkin, A.: Open boundary conditions for long-term integration of regional oceanic models, Ocean Model., 3, 1–20, doi:10.1016/S1463-5003(00)00013-5, 2001. 1075
Mogensen, K., Balmaseda, M. A., Weaver, A. T., Martin, M., and Vidard, A.: NEMOVAR: a variational data assimilation system for the NEMO ocean model, ECMWF Newsletter, 120, 17–22, 2009. 1085
Molcard, A., Griffa, A., and Ozgokmen, T. M.: Lagrangian data assimilation in multilayer primitive equation ocean models, J. Atmos. Ocean. Tech., 22, 70–83, 2004. 1088
Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Mon. Weather Rev., 129, 1494–1207, 2001. 1075, 1082
Primeau, F. W.: Multiple equilibria of a double-gyre ocean model with super-slip boundary conditions, J. Phys. Oceanogr., 28, 2130–2147, 1998. 1088
Pu, Z., Kalnay, E., Sela, J., and Szunyogh, I.: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method, Mon. Weather Rev., 125, 2479–2503, 1997. 1080
Reynolds, C. A. and Palmer, T. N.: Decaying singular vectors and their impact on analysis and forecast correction, J. Atmos. Sci., 55, 3005–3023, 1998. 1095, 1096
Roullet, G. and Madec, G.: Salt conservation, free surface, and varying levels: a new formulation for ocean general circulation models, J. Geophys. Res., 105, 23927–23942, 2000. 1087
Skamarock, W. C.: Evaluating mesoscale NWP models using kinetic energy spectra, Mon. Weather Rev., 132, 3019–3032, 2004. 1093
Stauffer, D. and Bao, J.-W.: Optimal determination of nudging coefficients using the adjoint equations, Tellus A, 45, 358–369, doi:10.1034/j.1600-0870.1993.t01-4-00003.x, 1993. 1076
Tenenhaus, M.: La régression PLS: Théorie et Pratique, 1st Edn., éditions Technip, Paris, France, 1998. 1083
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 Model., 13, 109–125, doi:10.1016/j.ocemod.2005.11.003, 2006. 1075
Verron, J.: Nudging satellite altimeter data into quasi-geostrophic ocean models, J. Geophys. Res.-Oceans, 97, 7479–7491, doi:10.1029/92JC00200, 1992. 1075

NPGD
1, 1073–1131, 2014

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, 2003. 1076

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, Ann. Glaciol., 54, 8–12, 2013. 1076

Weaver, A. T., Deltel, C., Machu, E., Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, Q. J. Roy. Meteorol. Soc., 131, 3605–3625, 2005. 1086

Willians, P. D.: A proposed modification to the Robert Asselin time filter, Mon. Weather Rev., 137, 2538–2546, 2009. 1087

Wold, H.: Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach, in: Perspectives in probability and statistics, Academic Press Inc., London, UK, 117–142, 1975. 1083

Wunsch, C.: The vertical partition of oceanic horizontal kinetic energy, J. Phys. Oceanogr., 27, 1770–1794, 1997. 1088

Zavala-Garay, J., Wilkin, J. L., and Arango, H. G.: Predictability of mesoscale variability in the east Australian current given strong-constraint data assimilation, J. Phys. Oceanogr., 42, 1402–1420, 2012. 1090

Zou, X., Navon, I. M., and Le Dimet, F. X.: An optimal nudging data assimilation scheme using parameter estimation, Q. J. Roy. Meteorol. Soc., 118, 1163–1186, doi:10.1002/qj.49711850808, 1992. 1076

NPGD
1, 1073–1131, 2014

Table 1. Summary of the experiments presented in Sect. 5.1. The symbol "_xxd" states for the length of the data assimilation window in days, i.e. ssh_10d_dd refers to an experiment assimilating SSH, using a 10 days DA window and default diffusion coefficients. In the Table "xx" may take the values: 2, 5, 10, 20 and 30.

	$K\,(s^{-1})$	$\nu_h^{u,v}\,(m^4\,s^{-1})$	$\nu_h^{t,s}\,(m^4\,s^{-1})$	convergence	Assim. Variables
ssh_xxd_dd	1.5×10^{-4}	-8×10^{10}	-4×10^{11}	0.5 %	SSH
ssh_xxd_rd	1.5×10^{-4}	-8×10^{9}	-4×10^{10}	0.5 %	SSH
ssh_xxd_rd_2it	1.5×10^{-4}	-8×10^{9}	-4×10^{10}	2 it	SSH

NPGD
1, 1073–1131, 2014

Table 2. Summary of the mean relative initial and final condition errors obtained from the DBFN experiments employing the reference diffusion and assimilating daily SSH observations. AW is the Assimilation window. e^0 and e^f are the mean initial and final errors, respectively.

AW	SSH		U		V		T	
	e^0	e^f	e^0	e^f	e^0	e^f	e^0	e^f
2d	0.0341	0.0613	0.2872	0.2891	0.3974	0.3815	0.0129	0.0129
5d	0.0369	0.0574	0.3138	0.3085	0.4503	0.4014	0.0137	0.0136
10d	0.0493	0.0649	0.3368	0.3279	0.5266	0.4175	0.0147	0.0142
20d	0.0760	0.0940	0.3558	0.3298	0.6649	0.4995	0.0159	0.0147
30d	0.1025	0.1312	0.3666	0.3473	0.7472	0.5840	0.0166	0.0149

NPGD
1, 1073–1131, 2014

Table 3. Summary of the mean relative initial and final condition errors obtained from the DBFN experiments employing a reduced diffusion and assimilating daily SSH observations. AW is the Assimilation window. For each AW the top lines represent the experiments considering only 2 iterations and the bottom line the experiments considering $\epsilon = 0.0005$. e^0 and e^f are the mean initial and final errors, respectively.

AW	SSH		U		V		T	
	e^0	e^f	e^0	e^f	e^0	e^f	e^0	e^f
2d	0.0505	0.0926	0.3142	0.3269	0.4272	0.4488	0.0118	0.0119
	0.0299	0.0595	0.2495	0.2590	0.3577	0.3755	0.0117	0.0117
5d	0.0412	0.0911	0.2958	0.3219	0.3993	0.4432	0.0116	0.0118
	0.0166	0.0373	0.2153	0.2194	0.2877	0.2927	0.0119	0.0119
10d	0.0401	0.0944	0.2854	0.3204	0.3904	0.4454	0.0116	0.0117
	0.0278	0.0412	0.2321	0.2243	0.3233	0.2801	0.0123	0.0122
20d	0.0608	0.1132	0.3130	0.3372	0.4644	0.4776	0.0119	0.0118
	0.0486	0.0685	0.2694	0.2517	0.4768	0.3724	0.0131	0.0126
30d	0.0815	0.1266	0.3403	0.3489	0.5643	0.5254	0.0124	0.0119
	0.0668	0.1078	0.2911	0.3035	0.5907	0.5153	0.0138	0.0130

NPGD
1, 1073–1131, 2014

Table 4. Summary of the mean error growth rate obtained from the DBFN experiments assimilating daily SSH observations. AW is the Assimilation window. For each AW the top lines represent the experiments considering the reference diffusion coefficient and the bottom lines the experiments considering a reduced diffusion coefficient.

AW	SSH	U	V	T
2d	1.3580	0.0926	−0.7921	0.0001
	1.4812	0.4711	0.8863	0.0023
5d	0.4115	−0.1055	−0.9784	−0.0036
	0.4152	0.0823	0.0985	−0.0001
10d	0.1554	−0.0884	−1.0919	−0.0050
	0.1336	−0.0782	−0.4319	−0.0014
20d	0.0901	−0.1302	−0.8269	−0.0058
	0.0998	−0.0885	−0.5218	−0.0026
30d	0.0954	−0.0644	−0.5439	−0.0054
	0.1367	0.0415	−0.2515	−0.0027

NPGD
1, 1073–1131, 2014

Table 5. Summary of the experiments presented in Sect. 5.2.

	convergence	Assim. Variables	Regressed Var.
ssh_10d_rd_uv	0.5%	daily full SSH field	UV
ssh_10d_rd_it2_uv	2 it	daily full SSH field	UV
ssh4_10d_rd_uv	0.5%	every 4 days full SSH field	UV
ssh4_10d_rd	0.5%	every 4 days full SSH field	–
ssh4_10d_rd_it2_uv	2 it	every 4 days full SSH field	UV
ssh4_10d_rd_it2	2 it	every 4 days full SSH field	–
DBFN + PLS	0.5%	every 4 days Jason1-like SSH	UVT
ONDG	direct	every 4 days Jason1-like SSH	UVT

NPGD
1, 1073–1131, 2014

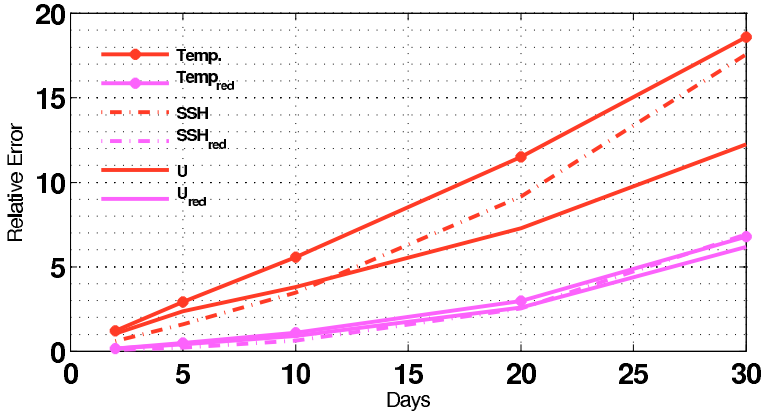

Figure 1. Errors of the initial condition after one forward-backward model integration perfectly initialized and without nudging. Red curves were obtained using the same diffusion coefficients used in the reference experiment ($\nu_h^{u,v} = -8 \times 10^{10}\,m^4\,s^{-1}$ and $\nu_h^{t,s} = -4 \times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

NPGD
1, 1073–1131, 2014

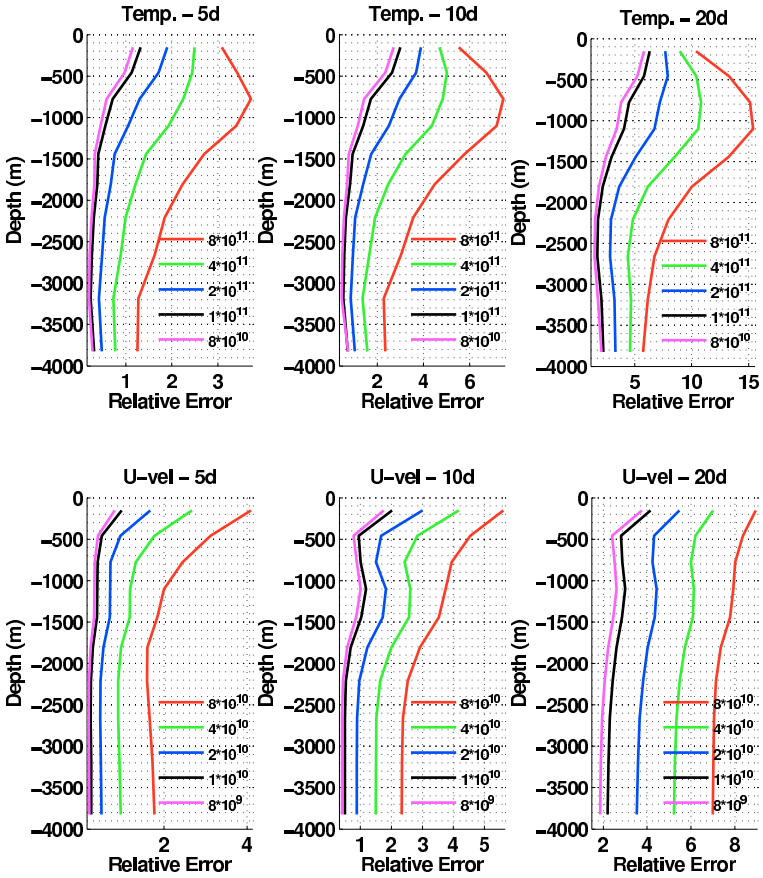

Figure 2. Vertical errors of the initial condition after one forward-backward model integration without nudging. Each color refers to an experiment performed using the diffusion coefficient indicated in the figures legend. Top panels: temperature errors; bottom panels: zonal velocity errors. The time window is indicated in the title of each figure.

NPGD
1, 1073–1131, 2014

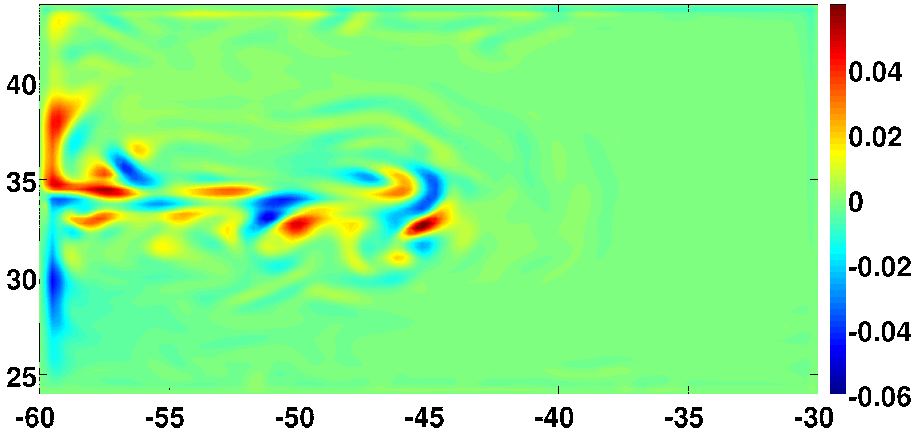

Figure 3. Sea level errors after one forward-backward model integration. The time window is of 10 days.

NPGD
1, 1073–1131, 2014

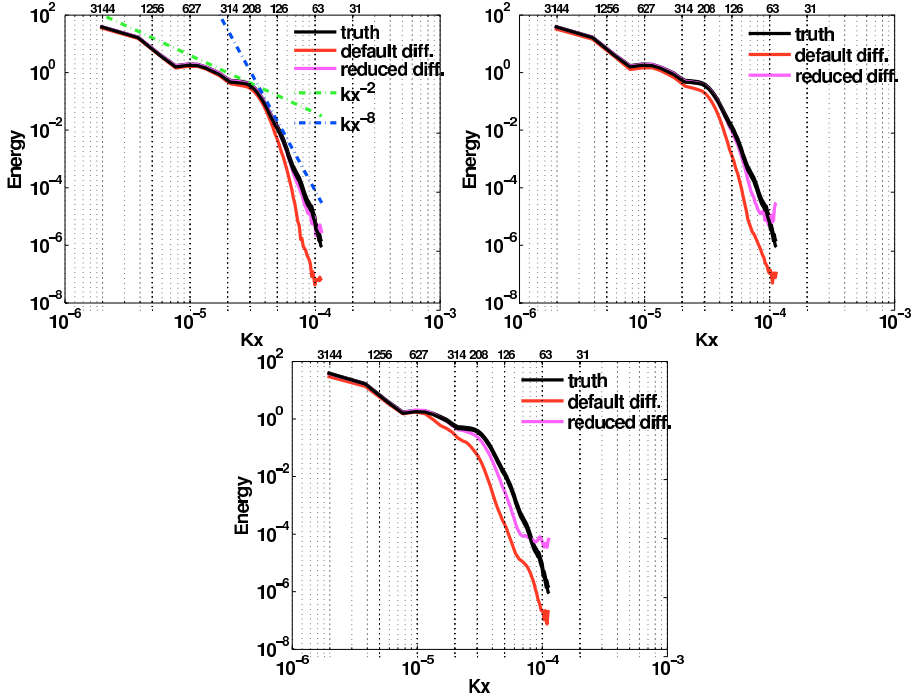

Figure 4. Kinetic energy mean power spectra calculated using the first layer velocity fields. Black curves represent the "true" initial condition power spectra; Red curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing the reference 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 panel: 5 days assimilation window. Top right panel: 10 days assimilation window. Bottom panel: 20 days assimilation window. In the bottom abscissa the ticklabels stand for longitudinal wave-number (rad m^{-1}) while in the top abscissa the ticklabels stand for the corresponding wavelengths in km units.

NPGD
1, 1073–1131, 2014

Numerical experiments with the DBFN

G. A. Ruggiero et al.

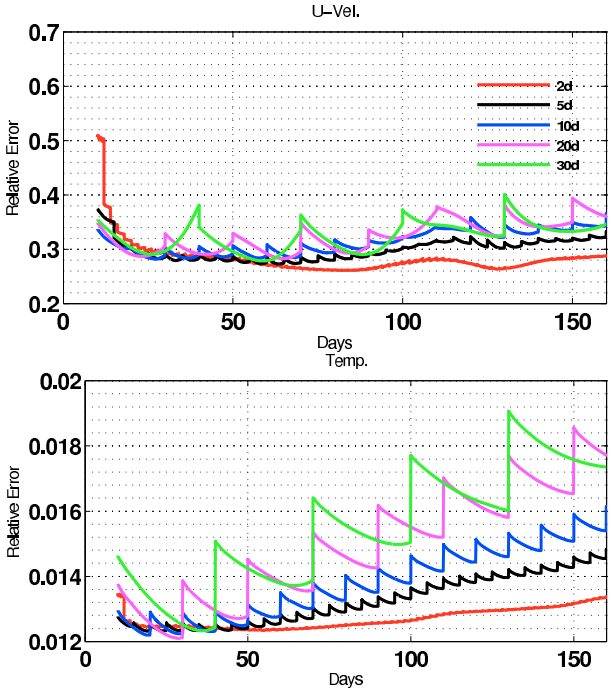

Figure 5. Relative errors of the zonal velocity (top panel) and the temperature (bottom panel) for the experiments listed in Table 1 which assimilates SSH using the convergence criterion and the reference diffusion coefficient.

NPGD
1, 1073–1131, 2014

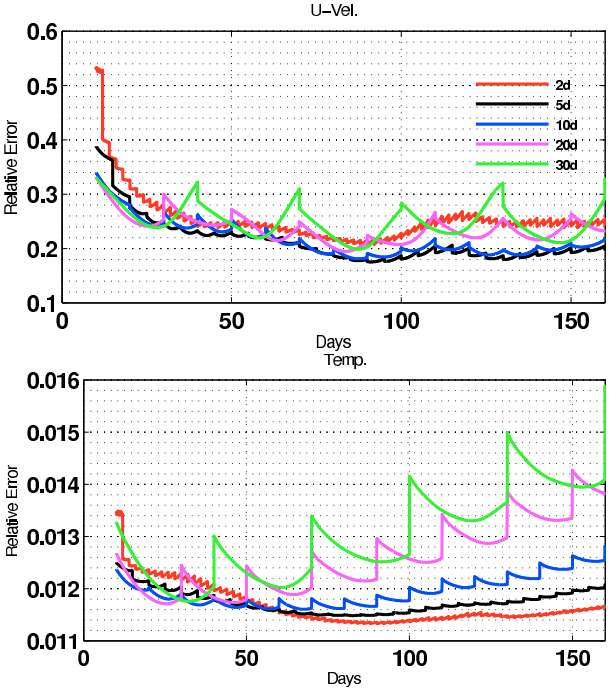

Figure 6. Relative errors of the zonal velocity (top panel) and the temperature (bottom panel) for the experiments listed in Table 1 which assimilates SSH using the convergence criterion and the reduced diffusion coefficient.

NPGD
1, 1073–1131, 2014

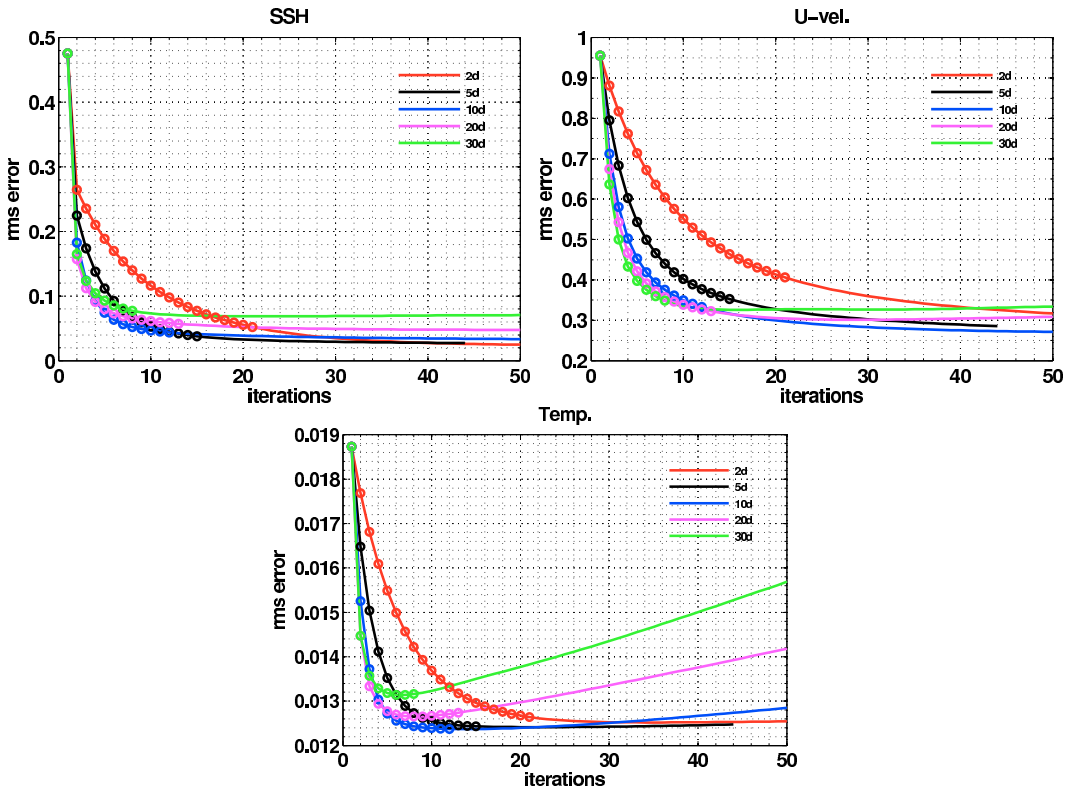

Figure 7. Variation of the initial condition relative errors with respect to the iterations for the experiment assimilating daily gridded SSH fields. The circles represent the results obtained using the standard convergence criterion, $\epsilon = 0.005$, and the continous lines obtained with a more restrictive criterion $\epsilon = 0.001$.

NPGD
1, 1073–1131, 2014

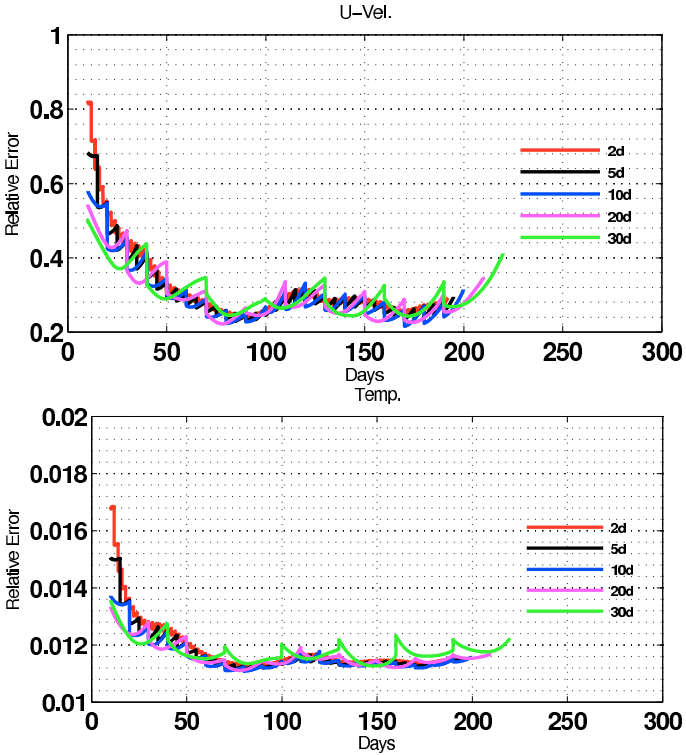

Figure 8. Relative errors of the zonal velocity and the temperature for the experiments listed in Table 1 which assimilates SSH using only two iterations. All experiments have used reduced diffusion coefficients.

NPGD
1, 1073–1131, 2014

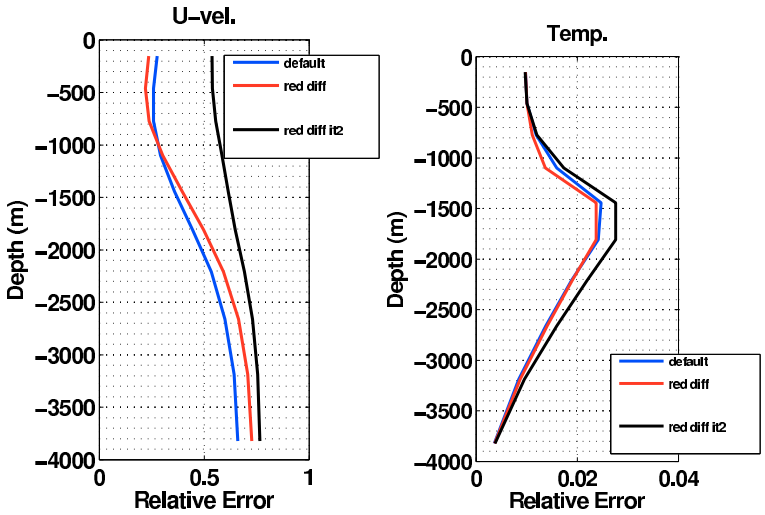

Figure 9. Vertical relative error of the zonal velocity and the temperature for the experiments ssh_10d_dd (default), ssh_10d_rd (red diff) and ssh_10d_rd_2it (red diff it2). The data refers to the identified initial conditions of the first assimilation cycle.

NPGD
1, 1073–1131, 2014

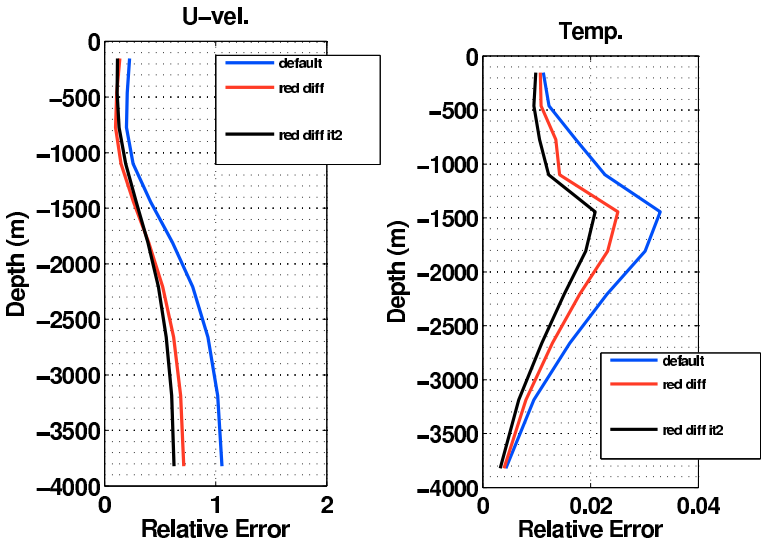

Figure 10. Vertical relative error of the zonal velocity and the temperature for the experiments ssh_10d_dd (default), ssh_10d_rd (red diff) and ssh_10d_rd_2it (red diff it2). The data refers to the identified initial conditions of the last assimilation cycle.

NPGD
1, 1073–1131, 2014

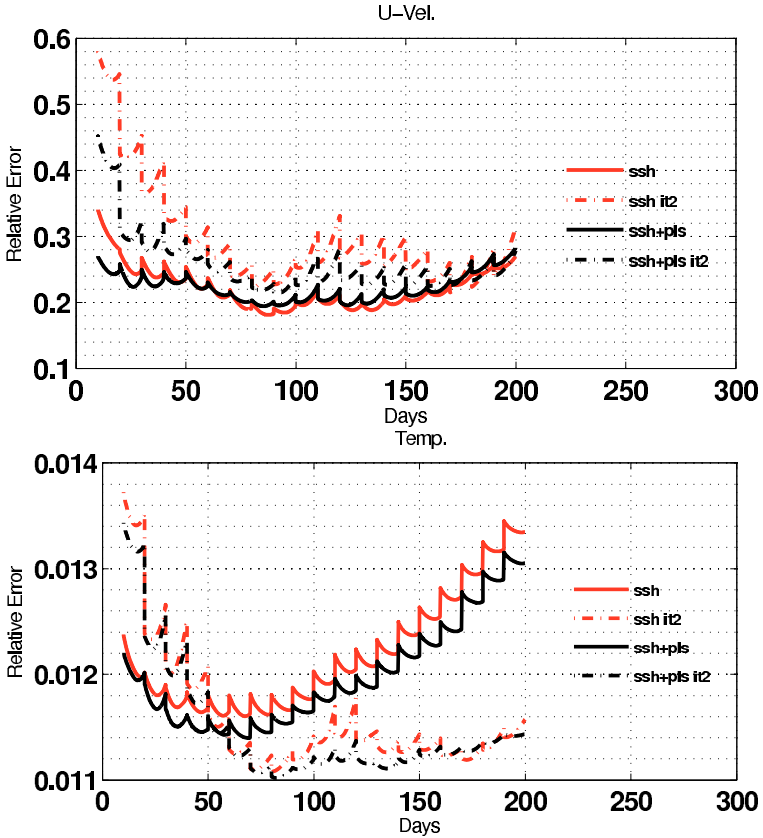

Figure 11. Relative errors of zonal velocity and temperature for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents but with the gain matrix **K** constructed using PLS regression (black curves).

NPGD
1, 1073–1131, 2014

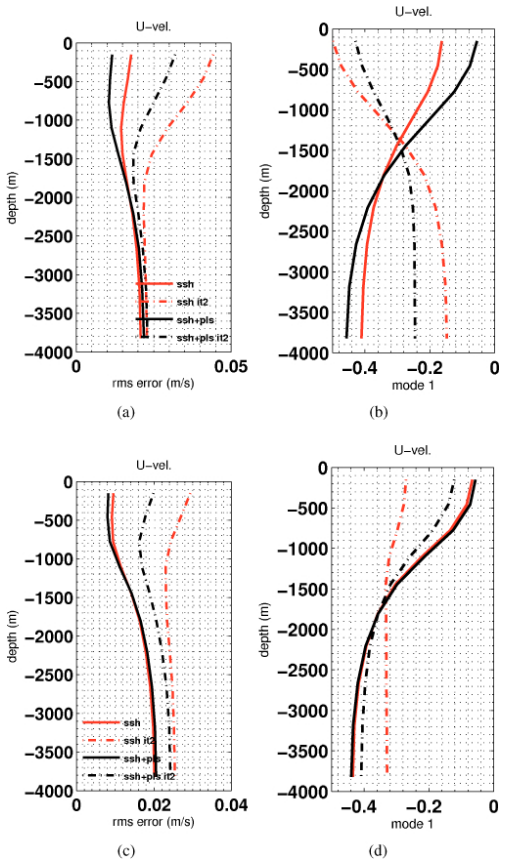

Figure 12. (a) and **(c)**: Root Mean Squared errors of zonal velocity for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents but with the matrix **K** constructed using PLS regression (black curves). **(b)** and **(d)**: first EOF mode calculated with the zonal velocity error. Top/bottom panels are results of the 1˚/130˚ day of the experiment.

NPGD
1, 1073–1131, 2014

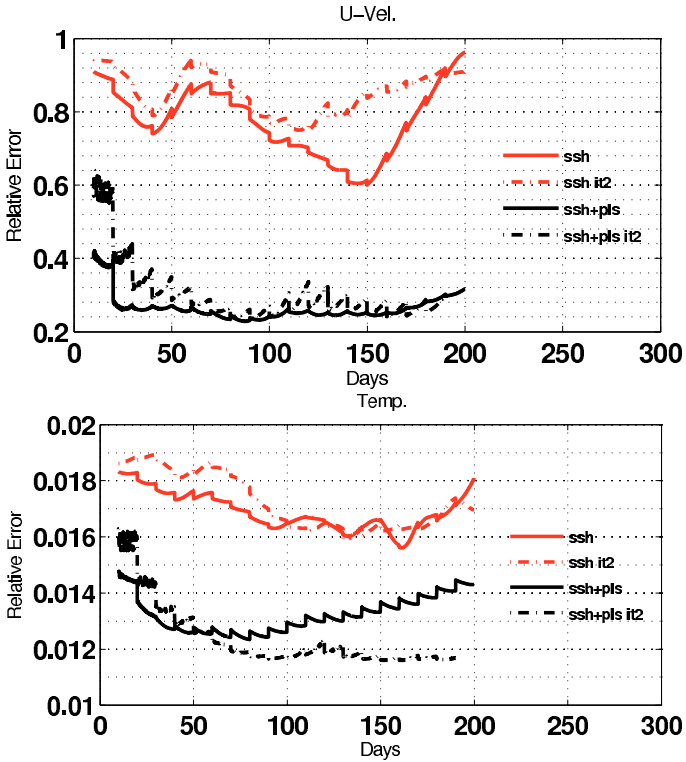

Figure 13. Relative errors of zonal velocity and temperature for the experiments ssh_10d_rd (red curve) and ssh_10d_rd_it2 (dashed red curve) and their equivalents but with the matrix **K** constructed using PLS regression (black curves).

NPGD
1, 1073–1131, 2014




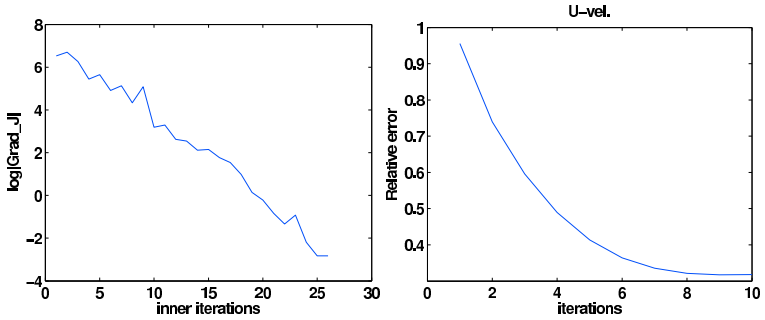

Figure 14. Figure shows the gradient of the cost function after each inner iteration (left) and the reduction of the relative error for zonal velocity for the experiment ssh4j_10d_rd_uvt (right).

NPGD
1, 1073–1131, 2014

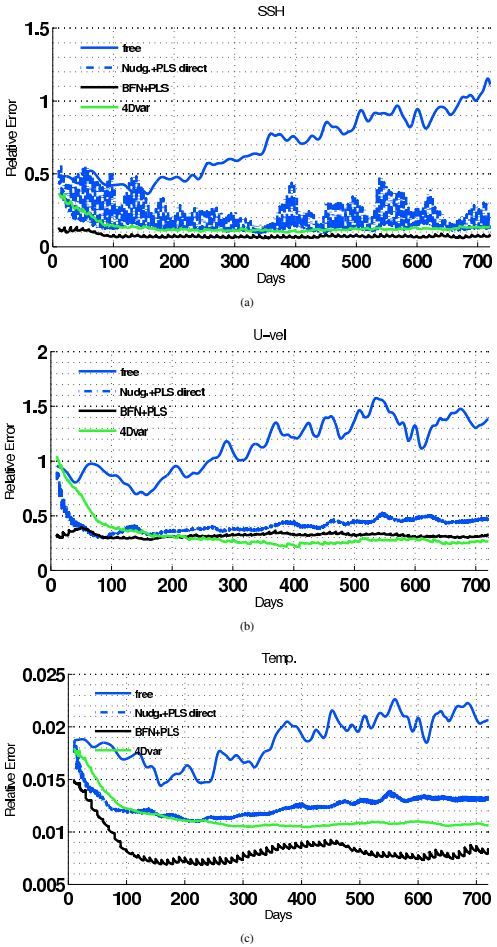

Figure 15. The figure shows errors of the SSH **(a)**, the zonal velocity **(b)** and the temperature **(c)**. Each curve correspond to a different experiment, see Table 5 for more details.

NPGD
1, 1073–1131, 2014

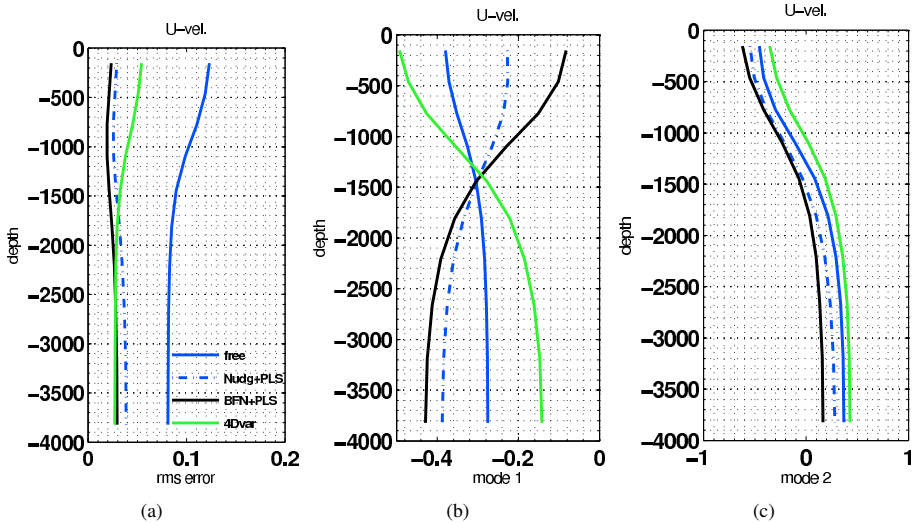

Figure 16. (a) RMS of vertical zonal velocity and first **(b)** and second **(c)** of error modes calculated using forecast from day 200 to day 720.

NPGD
1, 1073–1131, 2014