

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

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

¹Université de Nice Sophia-Antipolis/LJAD, Nice, France

² Université du Sud Toulon-Var, Aix-Marseille Université, CNRS/INSU, IRD, Mediterranean Institute of Oceanography (MIO), France

³Université Joseph Fourier/LGGE, Grenoble, France

⁴CNRS/LGGE, Grenoble, France

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

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

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

19 **1 Introduction**

20 The well-known Nudging method is based on the second Newton axiom and consists in adding a
21 forcing term in the right hand side of a given system in order to gently push the model toward a
22 prescribed value. The first appearance of nudging in the geophysical literature was in 1974 (Anthes,
23 1974). In this work the authors proposed the use of nudging to mitigate initialization problems in at-
24 mospheric models. However, a similar algorithm had already been developed by Luenberger (1966).
25 This algorithm has been called "Luenberger observer" or "asymptotic estimator", since under lin-
26 earity and observability hypothesis the estimator error converges to zero for time tending to infinity.
27 It is quite interesting to note that there is no mention of the Luenberger observer in the geophysical
28 literature except in the recent work of Auroux and Blum (2005). More recently, a comprehensive
29 study on the nudging method and its variants was produced by Blum et al. (2008) and Lakshmivar-
30 han and Lewis (2012).

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

42 The increasing availability of computing power has allowed to use more advanced data assim-
43 ilation methods. In general, these methods use information on the model statistics and observations
44 errors to weight the model-observations combination. Two of these methods that are widely used by
45 prediction centers are the ensemble Kalman filter- EnKF (Evensen, 1994) and its variations (Pham,
46 2001; Hunt et al., 2007), and the four dimensional variational method 4Dvar (Le Dimet and Tala-
47 grand, 1986; Courtier et al., 1994). For the first, the numerical costs are due to the propagation of the
48 ensemble, usually formed by tenths of members, to calculate the forecast. For the second, the costs
49 are due to the need of minimizing a cost function in a very large state space (10^8 variables). This
50 requires several iterations of the minimization algorithm, which involves several integrations of the
51 direct and adjoint models.

52 However, even with the growing interest in these complex techniques built on solid theoretical
53 arguments, nudging has not been left aside. Recent works have used nudging along with more
54 advanced methods such as Optimal interpolation (Clifford et al., 1997; Wang et al., 2013), EnKF
55 (Ballabrera-Poy et al., 2009; Bergemann and Reich, 2010; Lei et al., 2012; Luo and Hoteit, 2012),

56 4Dvar (Zou et al., 1992; Stauffer and Bao, 1993; Vidard et al., 2003; Abarbanel et al., 2010) or
57 particle filters (Luo and Hoteit, 2013; Lingala et al., 2013) to extract the best of each method. In
58 the particular case of the hybridization with the EnKF proposed by Lei et al. (2012), the resulting
59 algorithm takes the advantage of the dynamical propagation of the covariance matrix from the EnKF
60 and uses nudging to mitigate problems related to the intermittence of the sequential approach, which
61 among other things entails the possible discarding of some observations.

62 Recently, Auroux and Blum (2005) revisited the nudging method and proposed a new observer
63 called Back and Forth Nudging (BFN). The BFN consists in a sequence of forward and backward
64 model integrations, both of them using a feedback term to the observations, as in the direct nudg-
65 ing. The BFN integrates the direct model backwards in time avoiding the construction of the adjoint
66 and/or tangent linear models needed by 4DVar. Therefore, it uses only the fully non-linear model to
67 propagate information forward and backward in time. The nudging gain, which has an opposite sign
68 with respect to the forward case, has a double role: push the model toward observations and stabilize
69 the backward integration, which is especially important when the model is not reversible.

70 The BFN convergence was proved by Auroux and Blum (2005) for linear systems of ordinary
71 differential equations and full observations, by Ramdani et al. (2010) for reversible linear partial dif-
72 ferential equations (Wave and Schrödinger equations), by Donovan et al. (2010) and Leghtas et al.
73 (2011) for the reconstruction of quantum states and was studied by Auroux and Nodet (2012) for
74 non-linear transport equations. The BFN performance in numerical applications using a variety of
75 models, including non-reversible models such as a Shallow Water (SW) model (Auroux, 2009) and
76 a Multi-Layer Quasi-Geostrophic (LQG) model (Auroux and Blum, 2008), are very encouraging.
77 Moreover, by using a simple scalar gain, it produced results comparable to those obtained with
78 4DVar but with lower computational requirements (Auroux, 2009; Auroux et al., 2012).

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

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

93 Furthermore, the vertical ocean structure represented by NEMO is more complex than the verti-
94 cal ocean structure represented by the SW and LQG used by Auroux and Blum (2008) and Auroux
95 (2009). This is because the SW model has no vertical levels and the LQG was implemented with
96 only 3 layers, while in this article NEMO is configured with 11 vertical layers. In addition, NEMO
97 considers vertical diffusion processes, mostly ignored by the LQG model. Vertical diffusion plays an
98 important role in maintaining the ocean stratification and meridional overturning circulation, which
99 is directly related to the transport of heat in the ocean. Moreover from the practical point of view,
100 the diffusion/viscosity required to keep the NEMO simulations stable is by far greater than for the
101 SW or LQG at the same resolution.

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

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

112 **2 Data Assimilation Methods**

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

115 **2.1 The Back and Forth Nudging**

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

$$120 \quad \frac{d\mathbf{x}}{dt} = \mathcal{F}(\mathbf{x}) + \mathbf{K}(\mathbf{x}_{obs} - \mathcal{H}(\mathbf{x})) \quad (1)$$

121 where \mathbf{x} represents the state vector, \mathcal{F} is the model operator, \mathcal{H} is the observation operator allowing
122 one to compare the observations $\mathbf{x}_{obs}(t)$ to the corresponding system state $\mathcal{H}(\mathbf{x})$, and \mathbf{K} is the
123 nudging gain matrix. In this algorithm the model appears as a weak constraint. The feedback term
124 changes the dynamical equations and forces the state variables to fit the observations as well as pos-
125 sible.

126 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

127 of noise in the system, nudging is nothing else than the Luenberger observer (Luenberger, 1966). In
 128 this case, and assuming that the observability of the pair (\mathbf{F}, \mathbf{H}) holds, there is a class of possible
 129 values of \mathbf{K} that guarantees the estimator convergence when $t \rightarrow \infty$ (Gelb et al., 1974). This should
 130 be one possible explanation why nudging usually works quite well and the converged state is not
 131 strongly affected by the choice of \mathbf{K} . However, when constructing \mathbf{K} (which units is s^{-1}), the aim
 132 is to obtain an estimator response faster than the time scale of the studied processes.

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

138 The BFN novelty with respect to conventional nudging methods is the model integration back-
 139 ward in time. This allows to recover initial conditions as well as to use more than once the same
 140 observations set. Consequently, the BFN may be seen as a sub-optimal iterative smoother.

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

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

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

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

$$158 \quad \frac{\partial \mathbf{x}}{\partial t} = \mathcal{F}(\mathbf{x}) + \nu \Delta \mathbf{x}, \quad \text{for} \quad 0 < t < T, \quad (2)$$

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

$$162 \quad \frac{\partial \mathbf{x}_k}{\partial t} = \mathcal{F}(\mathbf{x}_k) + \nu \Delta \mathbf{x}_k + \mathbf{K}(\mathbf{x}_{obs} - \mathcal{H}(\mathbf{x}_k)) \quad (3)$$

163 $\mathbf{x}_k(0) = \tilde{\mathbf{x}}_{k-1}(0), \quad 0 < t < T,$

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

166
$$\frac{\partial \tilde{\mathbf{x}}_k}{\partial t} = \mathcal{F}(\tilde{\mathbf{x}}_k) - \nu \Delta \tilde{\mathbf{x}}_k - \mathbf{K}'(\mathbf{x}_{obs} - \mathcal{H}(\tilde{\mathbf{x}}_k)) \quad (4)$$

167 $\tilde{\mathbf{x}}_k(T) = \mathbf{x}_k(T), \quad T > t > 0.$

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

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

175
$$\frac{\|\mathbf{x}_k(t=0) - \mathbf{x}_{k-1}(t=0)\|}{\|\mathbf{x}_{k-1}(t=0)\|} \leq \epsilon, \quad (5)$$

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

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

184
$$\frac{\partial \mathbf{x}_\infty}{\partial t} = \mathcal{F}(\mathbf{x}_\infty) \quad (6)$$

185 while taking the difference between Eqs.(3) and (4) shows that \mathbf{x}_∞ satisfies the Poisson equation:

186
$$\Delta \mathbf{x}_\infty = -\frac{\mathbf{K}}{\nu}(\mathbf{x}_{obs} - \mathcal{H}(\mathbf{x}_\infty)) \quad (7)$$

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

190
$$\mathcal{J}(\mathbf{x}) = k < \mathbf{R}^{-1}(\mathbf{x}_{obs} - \mathbf{H}\mathbf{x}), (\mathbf{x}_{obs} - \mathbf{H}\mathbf{x}) > + \nu \int_{\Omega} \|\nabla \mathbf{x}\|^2 \quad (8)$$

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

196 The description of the used \mathbf{K} matrix is given in the Sect (5.1).

197 **2.2 Four Dimensional Variational Method - 4DVar**

198 Variational methods minimize a cost function that measures the distance between the estimated
 199 state and the available observations. Let us assume that observations are available at every instant
 200 $(t_i)_{1 \leq i \leq N}$. Given a first guess \mathbf{x}^b of the initial state, the 4DVar algorithm will find an optimal initial
 201 condition that minimizes the distance between the model trajectory and the observations in a given
 202 assimilation window. This optimal state is found by minimizing the following cost function:

$$203 \quad J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) \\ 204 \quad + \frac{1}{2} \sum_{i=0}^N (\mathcal{H}_i[\mathcal{M}_{0,i}(\mathbf{x}_0)] - \mathbf{y}_i)^T \mathbf{R}_i^{-1} (\mathcal{H}_i[\mathcal{M}_{0,i}(\mathbf{x}_0)] - \mathbf{y}_i) \quad (9)$$

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

208 The optimal initial state is found by solving:

$$209 \quad \nabla J(\mathbf{x}^a(t_0)) = 0 \quad (10)$$

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

212 If \mathcal{H} or \mathcal{M} are nonlinear, the solution of the problem is not unique, i.e. the functional (9) may
 213 have multiple local minima, and the minimization procedure may not stop at the global minimum. To
 214 overcome this problem, Courtier et al. (1994) proposed to solve a sequence of quadratic problems,
 215 expecting this sequence would converge to the solution of the problem given by (9) and (10). This
 216 algorithm is called the incremental 4Dvar. In this case, the cost function will not be minimized
 217 with respect to the initial state but with respect to an increment $\delta\mathbf{x}_0$ defined by $\mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0$. The
 218 operators \mathcal{H} or \mathcal{M} are linearized in a neighborhood of \mathbf{x}^b as:

$$219 \quad \mathcal{M}_{0,i}(\mathbf{x}^b + \delta\mathbf{x}_0) \approx \mathcal{M}_{0,i}(\mathbf{x}^b) + \mathbf{M}_{0,i}\delta\mathbf{x}_0 \quad \forall i \quad (11)$$

$$220 \quad \mathcal{H}_i(\mathbf{x}^b + \delta\mathbf{x}_0) \approx \mathcal{H}_i(\mathbf{x}^b) + \mathbf{H}_i\delta\mathbf{x}_0 \quad \forall i \quad (12)$$

221 and the new cost function is given by:

$$222 \quad J(\delta\mathbf{x}_0) = \frac{1}{2}\delta\mathbf{x}_0^T \mathbf{B}^{-1} \delta\mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^N (\mathbf{H}_i \mathbf{M}_{0,i} \delta\mathbf{x}_0 - \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \mathbf{M}_{0,i} \delta\mathbf{x}_0 - \mathbf{d}_i) \quad (13)$$

223 where $\mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\mathbf{x}_b))$ is called the innovation vector. It is possible that after some iterations
 224 of the minimizer the increments become too large and a new linearization of \mathcal{H} and \mathcal{M} should be
 225 done. This gives rise to what is called the inner loop and outer loop iterations. The algorithm
 226 implemented in NEMO, called NEMOVAR (Mogensen et al., 2009), uses this technics. It can be
 227 summarized as follows:

–Initialisation : $\mathbf{x}_0^0 = \mathbf{x}^b$
 –While $k \leq k_{max}$ or $\|\delta\mathbf{x}_0^{a,k}\| > \epsilon$ (**Outer Loop**)
 Do
 • $\mathbf{d}_i^k = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_{0,i}(\mathbf{x}_0^k))$
 • Search the $\delta\mathbf{x}_0^{a,k}$ that minimises (**Inner Loop**):

228

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

229 The description of the matrices \mathbf{B} and \mathbf{R} is given in the Sect (5.2).

230 **3 Ocean Model and Experimental set-up**

231 The ocean model used in this study is the ocean component of NEMO (Nucleus for European Mod-
 232 elling of the Ocean; Madec, 1996). This model is able to represent a wide range of ocean motions,
 233 from the basin scale up to the regional scale. Currently, it has been used in operational mode by the
 234 French Mercator Océan group (<http://www.mercator-ocean.fr>) and the European Center for Medium
 235 Range Weather Forecast (ECMWF).

236 The model solves six prognostic equations, namely the momentum balance, the hydrostatic equi-
 237 librium, the incompressibility equation, the heat and salt conservation equations and a nonlinear
 238 equation of state which couples the two tracers to the fluid fields. In this study, a linear free surface
 239 formulation is used along with the approach developed by Roullet and Madec (2000) to filter out the
 240 external gravity waves.

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

247 The double gyre configuration, extensively used to study jet instabilities (Chassignet and Gent,
 248 1991; Primeau, 1998; Chang et al., 2001), meso and submeso-scale dynamics (Levy et al., 2010)
 249 and data assimilation methods (Molcard et al., 2004; Krysta et al., 2011; Cosme et al., 2010), is used
 250 for the present study. The double gyre configuration simulates the ocean middle latitude dynamics
 251 and has the advantage of being simple, when compared to real applications, but still considering full
 252 dynamics and thermodynamics.

253 In our experiments we use a homogeneous horizontal grid with a 25km resolution and a verti-

254 cal resolution ranging from 100m near the upper surface to 500m near the bottom. The bottom
 255 topography is flat and the lateral boundaries are closed and frictionless. The only forcing term
 256 considered is a constant wind stress of the form $\tau = (\tau_0 \cos(\frac{2\pi(y-y_0)}{L}), 0)$, where y is the lati-
 257 tude geographic coordinate with $y_0 = 24^\circ$ and $24^\circ \leq y \leq 44^\circ$, $L = 20^\circ$ and $\tau_0 = -0.1 \text{ N/m}^2$. Hori-
 258 zontal diffusion/viscosity are modeled by a bilaplacian operator meanwhile a laplacian operator is
 259 used in the vertical. They all use constant coefficients in time and space: $\nu_h^{u,v} = -8 \times 10^{10} \text{ m}^4/\text{s}$
 260 and $\nu_v^{u,v} = 1.2 \times 10^{-4} \text{ m}^2/\text{s}$ for the momentum equations and $\nu_h^{t,s} = -4 \times 10^{11} \text{ m}^4/\text{s}$ and $\nu_v^{t,s} =$
 261 $1.2 \times 10^{-5} \text{ m}^2/\text{s}$ for temperature and salinity. The initial condition is similar to that used by Chas-
 262 signet and Gent (1991) and consists of a homogeneous salinity field of 35psu and a temperature field
 263 created to provide a stratification which has a first baroclinic deformation radius of 44.7km. Velocity
 264 and sea surface height (SSH) fields are initially set to zero.

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

272 4 The backward integration without Nudging: Practical aspects

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

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

$$284 R_{\text{err}} = \frac{\|\mathbf{x}(0) - \tilde{\mathbf{x}}(0)\|}{\langle \|\mathbf{x}(t + \Delta t) - \mathbf{x}(t)\| \rangle} \quad (15)$$

285 where $\Delta t = 1 \text{ day}$ and the brackets represent the empirical mean.

286 Figure 1 shows the global error, R_{err} , for different window sizes. The errors grow linearly with
 287 the window size for all variables. Temperature is the most affected variable, followed by sea level
 288 and velocities. Temperature errors exceed 18 times a typical one-day variation for the 30 days exper-

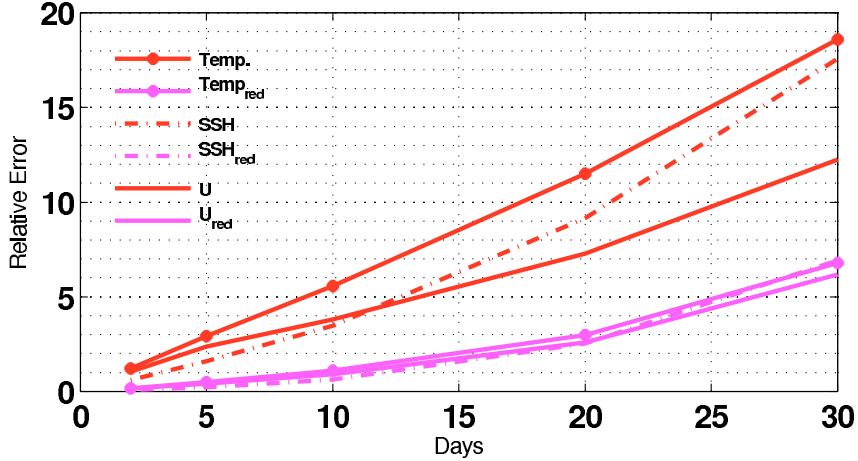


Fig. 1. Errors on the initial condition after one forward-backward model integration perfectly initialized and without nudging. Red curves were obtained using the same diffusion coefficients as in the reference experiment ($\nu_h^{u,v} = -8 \times 10^{10} m^4/s$ and $\nu_h^{t,s} = -4 \times 10^{11} m^4/s$)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

289 iment and 1.2 times for the 2 days. The use of reduced diffusion/viscosity coefficients reduces the
 290 errors to 6.8 and 0.16 times the one-day variation for 30 and 2 days experiments, respectively. Ve-
 291 locities errors were reduced by 50% for 30 days and 85% for 2 days, while ssh errors were reduced
 292 by 60% and 88% for 30 and 2 days, respectively.

293 As shown on Fig. 2 velocity and temperature errors are depth-dependent. Whereas for velocity
 294 they are larger at the surface and decrease with depth, for temperature they are larger in the ther-
 295 mocline. In the cases for which the forward-backward integrations use the same diffusion/viscosity
 296 coefficients as in the reference simulation, the temperature errors at thermocline depths exceed 3
 297 times the typical one day variation for the 5 days experiments and reaches 15 times for 20 days ex-
 298 periments. Considering the velocities, errors are proportional to 4 one-day variations for the 5 days
 299 experiment and to 8 one-day variations for the 20 days experiments. For time windows of 10, 20 and
 300 30 days, velocities at the thermocline depths start to be influenced by temperature errors.

301 Reduction of the diffusion/viscosity coefficients greatly reduced the errors especially in the ther-
 302 mocline for the temperature and at the surface for the velocity. It can be noted that when the diffusion
 303 coefficient is decreased the errors converge to a limit. This limit changes with respect to the window
 304 length and should be related to the diffusion required to stabilize the numerical method, which is of
 305 second order in our case, and hence oscillatory. Therefore, there is a compromise between the errors
 306 induced by the extra diffusion and errors due to spurious oscillations.

307 Numerical errors were assessed by changing the model time step from 900s to 90s. The resulting
 308 errors (not shown) do not change, suggesting that the errors induced by the diffusion are domi-

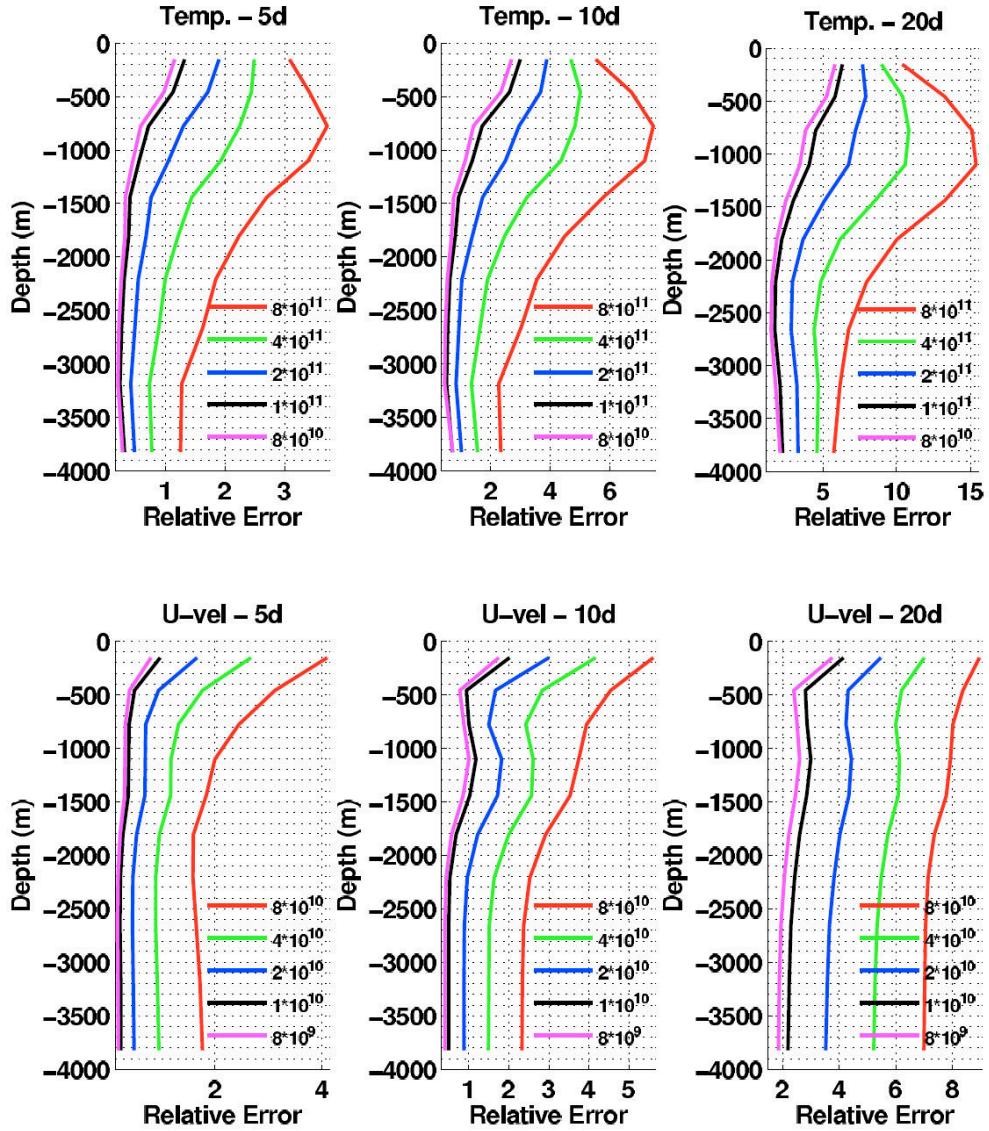


Fig. 2. Vertical profiles of relative errors on the initial condition after one forward-backward model integration without nudging. Each color refers to an experiment performed using the diffusion coefficient indicated in the figures legend. Red curves were obtained using the same diffusion coefficients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

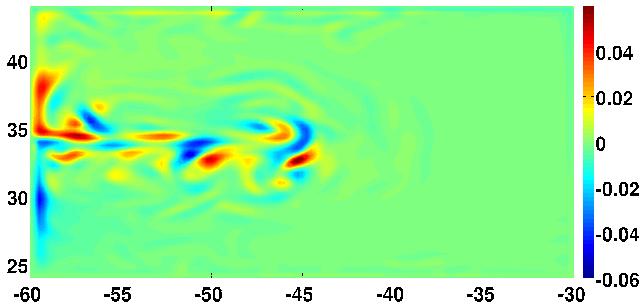


Fig. 3. Sea level errors after one forward-backward model integration. The time window is of 10 days.

309 nant. On the one hand, this is important because the complete rewriting of the model's code can be
 310 difficult, similarly to the adjoint model programming used by the 4Dvar, but on the other hand if
 311 the assimilation cannot control the diffusion errors it may represent a fundamental problem of the
 312 method when it is applied to non-reversible geophysical systems such as the ocean.

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

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

324 The longer the time window the greater the portion of the spectra affected. For the experiment
 325 employing the reference diffusion coefficient, the divergence between the true spectra and the spec-
 326 tra obtained from the backward integration is observed at 126, 314 and 627km for 5, 10 and 20 days
 327 experiments, while for the experiments considering a reduced diffusion coefficient there is almost
 328 no differences for the 5 days experiment, and the divergence is observed at 126 and 314km for the
 329 10 and 20 days experiments. If on the one hand using the reduced diffusion helps to keep the en-
 330 ergy distribution coherent with the true distribution, on the other hand it creates noise in the range
 331 of 126km to 25km. This confirms that there is a trade-off between the errors due to the excessive
 332 smoothing and the errors due to high frequency numerical modes.

333 In this section we have seen that there are large backward-errors induced by over-diffusion.
 334 Therefore, short time windows with reduced diffusion coefficients would be preferable to be used
 335 in DA experiments. Two regions have to be cautiously analyzed: the surface and the thermocline.

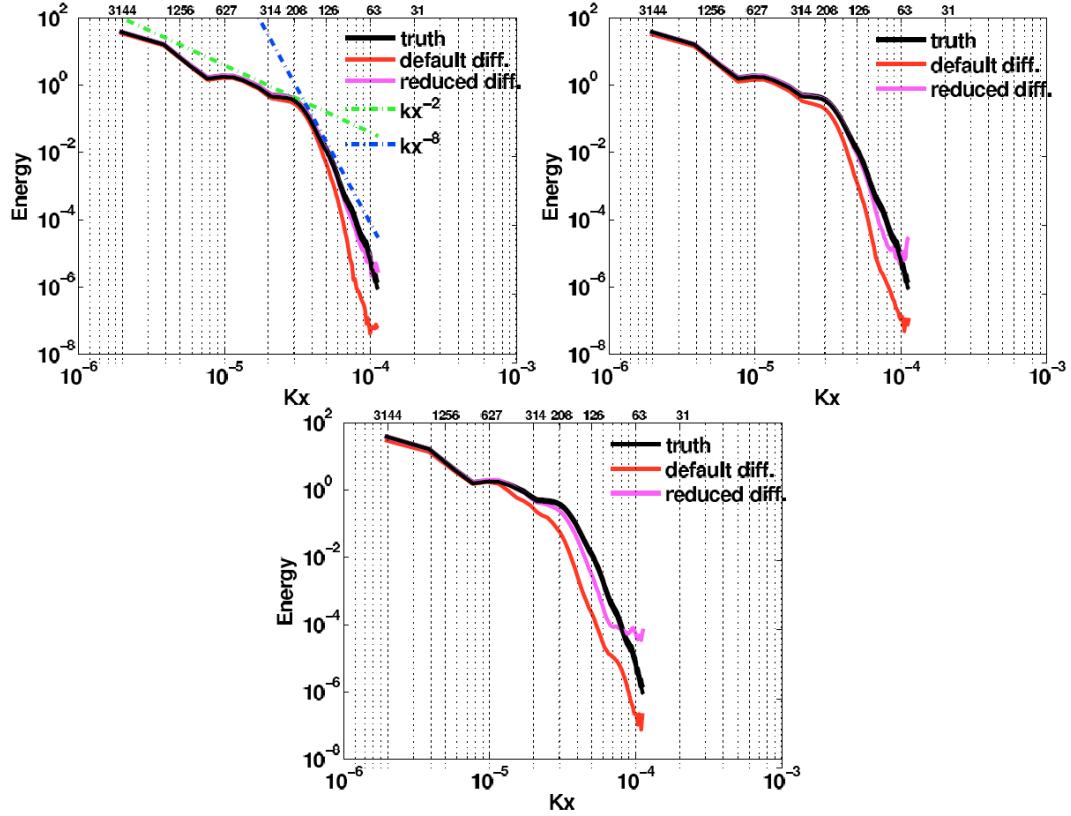


Fig. 4. Kinetic energy mean power spectra calculated using the first layer velocity fields. Black curves represent the “true” initial condition power spectra; Red curves represent the power spectra calculated after one forward-backward iteration without the nudging term and employing the reference diffusion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

336 Surface layers are prone to feature errors due to their role on the wind energy dissipation while at
337 the thermocline strong density gradients contribute to high diffusion rates.

338 **5 Data Assimilation experiments**

339 **5.1 Prescription of the DBFN gain**

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

344
$$\delta\mathbf{y} = \Theta(\mathbf{x}^{obs} - \mathbf{y}^b) \quad (16)$$

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

347
$$\Theta = \frac{\sigma_m^2}{\sigma_m^2 + \gamma\sigma_o^2} \quad (17)$$

348 where σ_m^2 is the mean spatial value of SSH variance calculated from the free model run, σ_o^2 is the
349 observation error variance and γ is a parameter used to adjust the variance of the observation errors.
350 When $\gamma = 1$ the Eq.(17) for the weight Θ has the same form of the scalar Kalman gain (Gelb et al.,
351 1974). For values greater than one, γ is an inflation factor, i.e. it increases the variance of the
352 observation errors resulting in more weight given to the model in the Eq.(16).

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

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

365
$$\delta\mathbf{x} = \hat{\mathbf{B}}^{PLS} \delta\mathbf{y} \quad (18)$$

366 where $\hat{\mathbf{B}}^{PLS}$ is the Partial Least Squares (PLS) regression coefficients which are described below.
367 It is worth noting that in Sect 6 we also apply this update scheme to an ordinary direct nudging

368 experiment. In this case γ is equal to one.

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

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

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

$$382 \quad \mathbf{T} = \mathbf{Y}\mathbf{W}, \quad (19)$$

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

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

$$386 \quad \mathbf{w}_i = \operatorname{argmax}_{\mathbf{w}} \{ \operatorname{cov}(\mathbf{Y}\mathbf{w}, \mathbf{X})^2 \} \quad (20)$$

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

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

$$389 \quad \hat{\mathbf{B}}^{PLS} = \mathbf{W}(\mathbf{W}^T \mathbf{Y}^T \mathbf{Y} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y}^T \mathbf{X} \quad (21)$$

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

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

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

400 5.2 The 4Dvar background term configuration

The 4Dvar considers a background term of the form:

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

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

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

$$417 \quad \mathbf{temp} = \mathbf{temp} \tag{22}$$

$$418 \quad \mathbf{salt} = \mathbf{salt}_B + \mathbf{salt}_U = \mathbf{G}_{\mathbf{salt}, \mathbf{temp}}(\mathbf{temp}) + \mathbf{salt}_U \tag{23}$$

$$419 \quad \eta = \eta_B + \eta_U = \mathbf{G}_{\eta, \rho}(\rho) + \eta_U \tag{24}$$

$$420 \quad \mathbf{u} = \mathbf{u}_B + \mathbf{u}_U = \mathbf{G}_{\mathbf{u}, \rho}(\rho) + \mathbf{u}_U \tag{25}$$

$$421 \quad \mathbf{v} = \mathbf{v}_B + \mathbf{v}_U = \mathbf{G}_{\mathbf{v}, \rho}(\rho) + \mathbf{v}_U \tag{26}$$

$$422 \quad \mathbf{p} = \mathbf{p}_B + \mathbf{p}_U = \mathbf{G}_{\mathbf{p}, \rho}(\rho) + \mathbf{p}_U \tag{27}$$

423 where

$$424 \quad \rho = \mathbf{G}_{\rho, \mathbf{temp}}(\mathbf{temp}) + \mathbf{G}_{\rho, \mathbf{salt}}(\mathbf{salt}) \tag{28}$$

$$425 \quad \mathbf{p} = \mathbf{G}_{\mathbf{p}, \rho}(\rho) + \mathbf{G}_{\mathbf{p}, \eta}(\eta) \tag{29}$$

with ρ the density and \mathbf{p} the pressure.

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

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

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

432

433 5.3 Assimilation cycle

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

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

440 The length of the Data Assimilation window (DAw) used in the reference experiments (Sect 6.1) is
441 10 days. For the sensitivity experiments presented in the Sect 6.2 the lengths of the the assimilation
442 window are 5 days and 30 days.

443 5.4 Observation network

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

447 The data assimilation problem we proposed to solve is to recover the full model state at the begin-
448 ning of the assimilation window. The model state space is composed of five variables: sea surface
449 height (η), meridional and zonal velocities (u and v), temperature and salinity ($temp$ and $salt$).
450 Since we have a horizontal mesh of size 81×121 and 11 vertical layers the total size of the state
451 space is 441045. Therefore, the problem is undetermined, since the observations represent only a
452 1.1% of the total state space. This means that the background term, and accordingly the B matrix
453 for the 4Dvar and the regression model \hat{B}^{PLS} for the DBFN, have quite a strong importance on the
454 method performances since they project the increments of the observed variables onto the numerous
455 non-observed variables.

456 To study at which extent the results are depend on the amount of assimilated observations and on
457 the first guess, in Sect 6.2.2 two additional experiments assimilating complete daily fields of SSH
458 are conducted: one using the same first guess of the experiments of Sect 6.1, and another using a
459 perturbed initial condition. In despite of the fact that the problem continues to be underestimated,
460 in this case the SSH analysis is no more dependent on the SSH spatial covariance, and the unstable

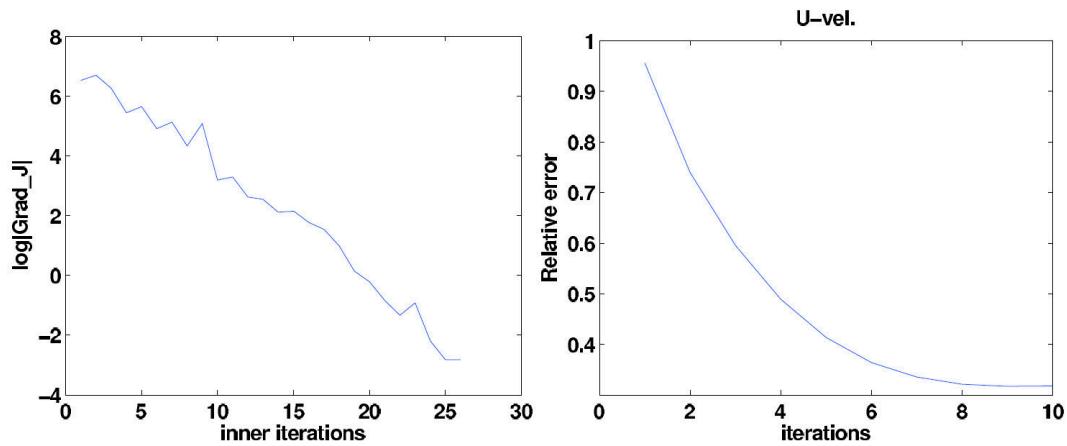


Fig. 5. Figure shows the gradient of the cost function after each inner iteration (left) and the reduction of the relative error for zonal velocity for the DBFN experiment (right).

461 modes associated with the SSH dynamics are certainly observed. The analysis produced for the
 462 other state vector variables remains dependent on the matrices \mathbf{B} for the 4Dvar case and $\hat{\mathbf{B}}^{PLS}$ for
 463 the DBFN case.

464 6 Data Assimilation Results

465 6.1 Reference experiment

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

472 First the minimization performance of the 4Dvar implementation is analysed. Figure 5 shows the
 473 reduction of the cost function gradient for the 4Dvar and the reduction of the relative error of the
 474 zonal velocity for the DBFN, both of them for the first assimilation cycle. 4Dvar takes 26 iterations
 475 to approximately achieve the optimality condition $\nabla J = 0$. This represents 3 times the number of it-
 476 erations required by the DBFN to converge, i.e., after which the errors cease to decrease. Moreover,
 477 the 4Dvar numerical cost is more than 3 times the DBFN cost since one execution of the adjoint
 478 model costs four times the cost of the direct model in terms of CPU time.

479 We note that the minimum error for the DBFN is reached after 9 iterations. This is quite consis-
 480 tent with our choice $\gamma = 18$, since theoretically it allows the use of the same set of observations for
 481 18 times.

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

487

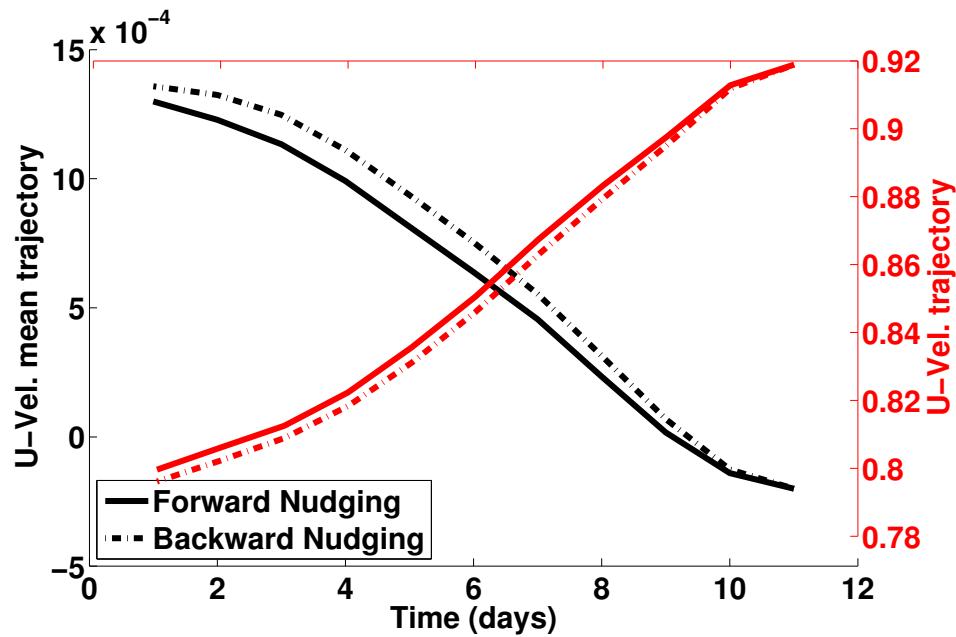


Fig. 6. Black curves represent the forward and backward surface zonal velocity mean trajectories at convergence and red curves the forward and backward surface zonal velocity trajectories at convergence for a point located at 34° North and 52.6° West, which is located on the unstable jet.

488 Figure 7 shows the root mean squared (rms) error for the control experiment (without assim-
 489 ulation), the experiment using the direct nudging with PLS regression (ONDG), the DBFN and the
 490 4Dvar. The DBFN errors for the velocity and SSH converge to their asymptotic values after the
 491 first assimilation cycle while for ONDG and 4Dvar errors stop decreasing after 100 and 200 days,
 492 respectively. This is a benefit of the iterations performed by the DBFN when model and data are
 493 quite different. Among the experiments conducted, the DBFN produced the smallest errors for all
 494 variables, except for the zonal velocity, for which the 4Dvar has slightly smaller errors. The ONDG
 495 also showed good performance, but with errors larger than the DBFN and 4Dvar errors.

496 With respect to the vertical error (Fig. 8), the DBFN and the ONDG performed better for the
 497 upper ocean than 4Dvar. Clearly, the PLS also corrects the deep ocean velocity, but less accurately
 498 than 4Dvar. The first error mode is the barotropic one, i.e. it has the same sign over all depths, and
 499 accounts for 97% of the error variability for 4Dvar, 96% and 93% for DBFN and ONDG, respec-

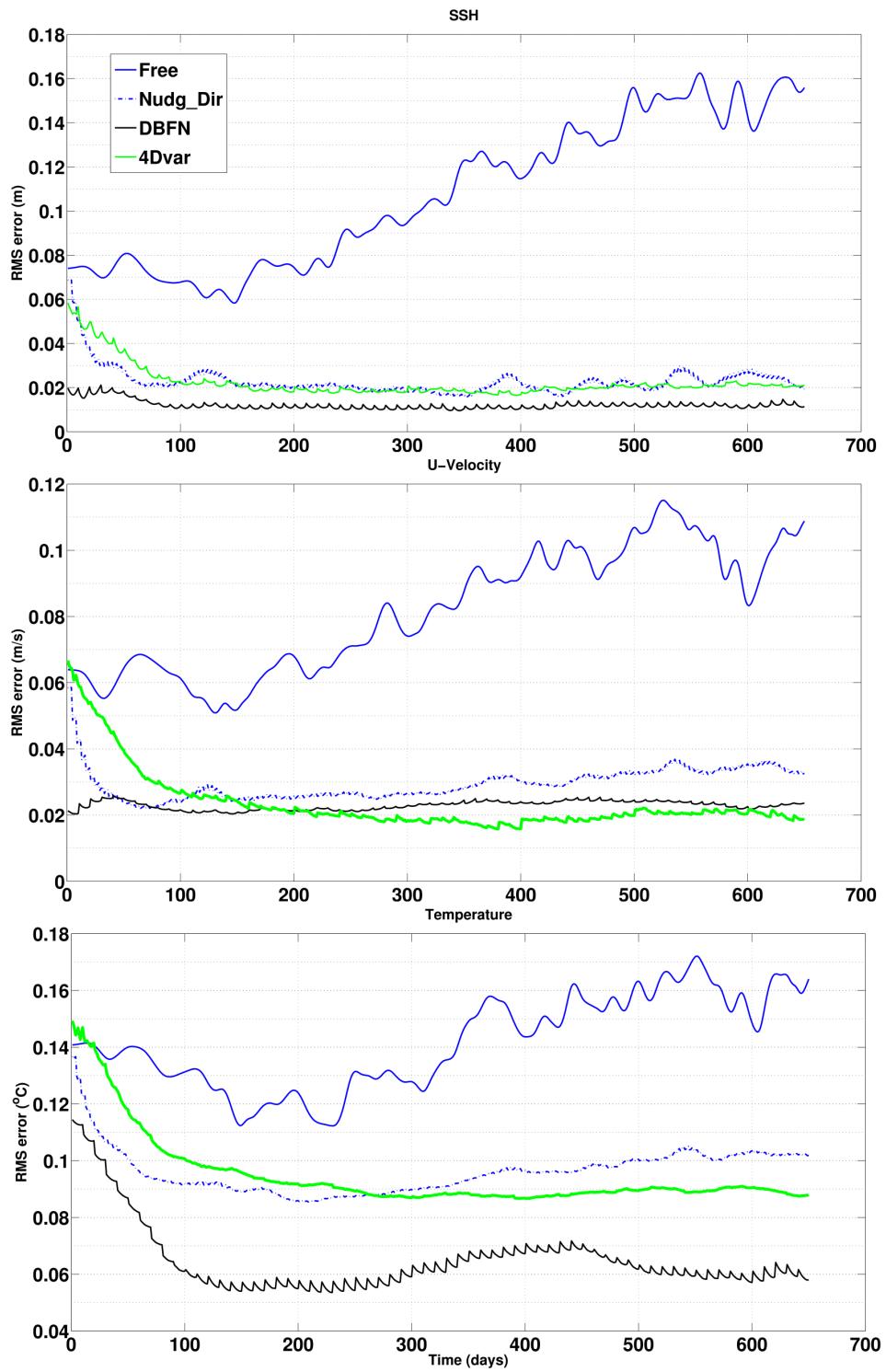


Fig. 7. The figure shows errors of the SSH (top panel), the zonal velocity (middle panel) and the temperature (bottom panel).

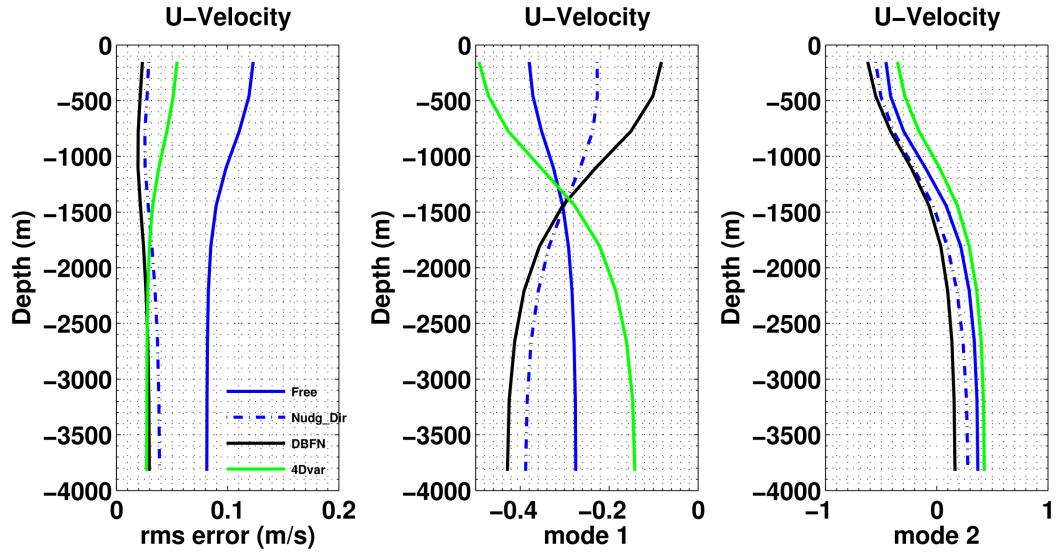


Fig. 8. Vertical profiles of rms error in zonal velocity (Left panel) and first (middle panel) and second (right panel) eof error modes calculated using forecast from day 200 to day 720.

500 tively. Although the first mode is the barotropic one for all methods, the 4Dvar barotropic mode of
 501 error is out of phase with respect to the PLS barotropic mode. This reflects the better performance
 502 of the 4Dvar for the deep ocean and the better performance of the DBFN and ONDG for the upper
 503 ocean.

504 The second mode, which accounts for almost all the remaining variability, has a sign inversion
 505 with depth and is found especially over the main axis of the jet. In this region the deep ocean veloc-
 506 ities are overestimated due to spurious covariances between the SSH and the deep ocean velocities.

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

517

518 6.2 Sensitivity experiments

519 6.2.1 Length of the DAw

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

524 Figure 9 shows the evolution of the rms errors for the zonal velocity and temperature during the
525 DBFN iterations over the first assimilation cycle, for three DAw (including the ten day-window used
526 previously). When considering only one iteration, the best results were obtained with the 30 days-
527 window experiment. This is a consequence of the asymptotic character of the Nudging method: the
528 longer the assimilation window, the more observations accounted for, the smaller the error. This
529 changes when several iterations are considered. The observed divergence for the 30 days-window is
530 due to the errors induced by the over-diffusion that induce great increments, which by their nature,
531 are not modelled by the ensemble of model states used to construct the regression model.

532

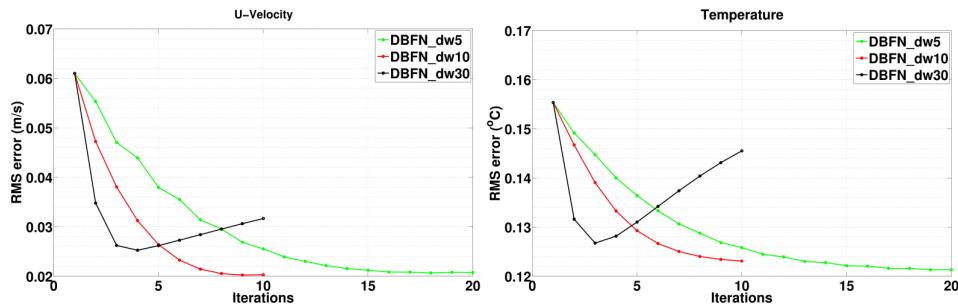


Fig. 9. Evolution of the rms errors for the zonal velocity and temperature during the DBFN iterations over the first assimilation cycle, for three DAw: 5, 10 and 30 days.

533 Figure 10 shows the rms error for the DBFN and 4Dvar experiments for three DAw: 5, 10 and
534 30 days. The methods exhibited comparable performance depending on the length of the DAw. For
535 the DBFN the 5 and 10 days DAw provided better results than the 30 days window, while for the
536 4Dvar the 30 days window provided the best estimation in terms of rms error. The DBFN and 4Dvar
537 experiments using the 30 and 5 days DAw, respectively, failed to identify the initial conditions since
538 their SSH rms errors are greater than the observation error standard deviation. The poor performance
539 of the 4Dvar for the 5 days DAw is related to spurious increments due to the fact that in one assim-
540 ilation window there is only one set of observation available. If this set is at the end of the window
541 this can complicate the minimization process and the iterations may stop before convergence.

542 Figure 11 shows the time evolution of vertical profiles of horizontally layer-wise averaged rms

543 error of zonal velocities for the DBFN and 4Dvar experiments. The 4Dvar profits of the longer DAw
544 to spread the observation to the 3-dimensional variables. This is done by the iterations of the direct
545 model and by the B matrix. For the DBFN experiments, after one year of data assimilation the
546 errors in the deep ocean start to grow. This is due to the high variance of the PLS estimator for deep
547 layers. The problem becomes more evident on the second year because at this stage the observa-
548 tions are farther from the model states used to construct the regression coefficients. Therefore, this
549 mean that this behavior is not intrinsic to the DBFN algorithm and its diffusive aspects, but due to
550 our implementation. Ideally, the regression model should evolve in time, similarly to the Kalman
551 Filter scheme. The 4Dvar has good performance at the deep ocean thanks to the use of a vertical
552 localization with a length scale of 1500m.

553

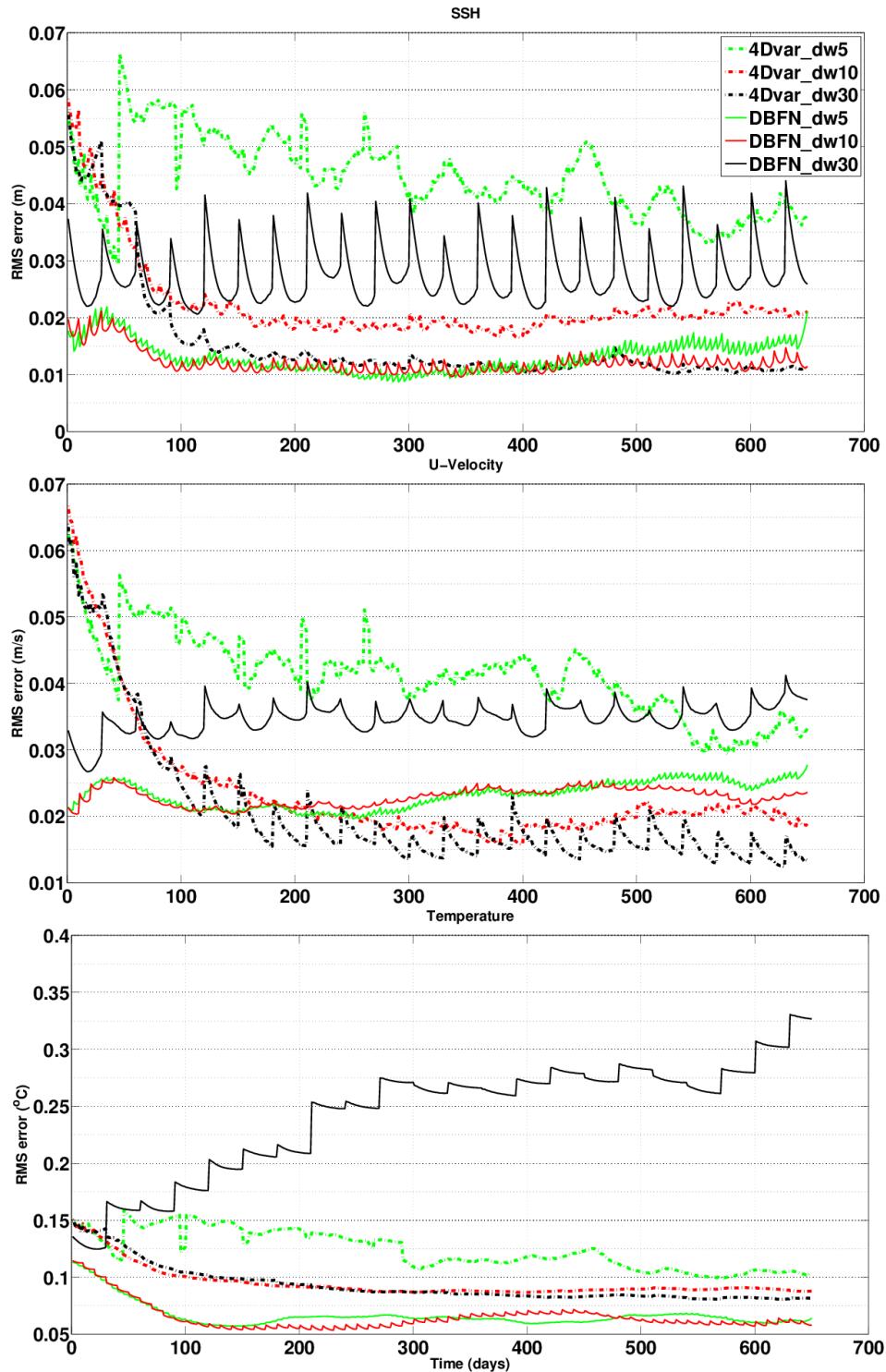


Fig. 10. RMS errors on SSH (top panel), zonal velocity (middle panel) and temperature (bottom panel) from DBFN and 4Dvar experiments with DA_W of 5, 10 and 30 days.

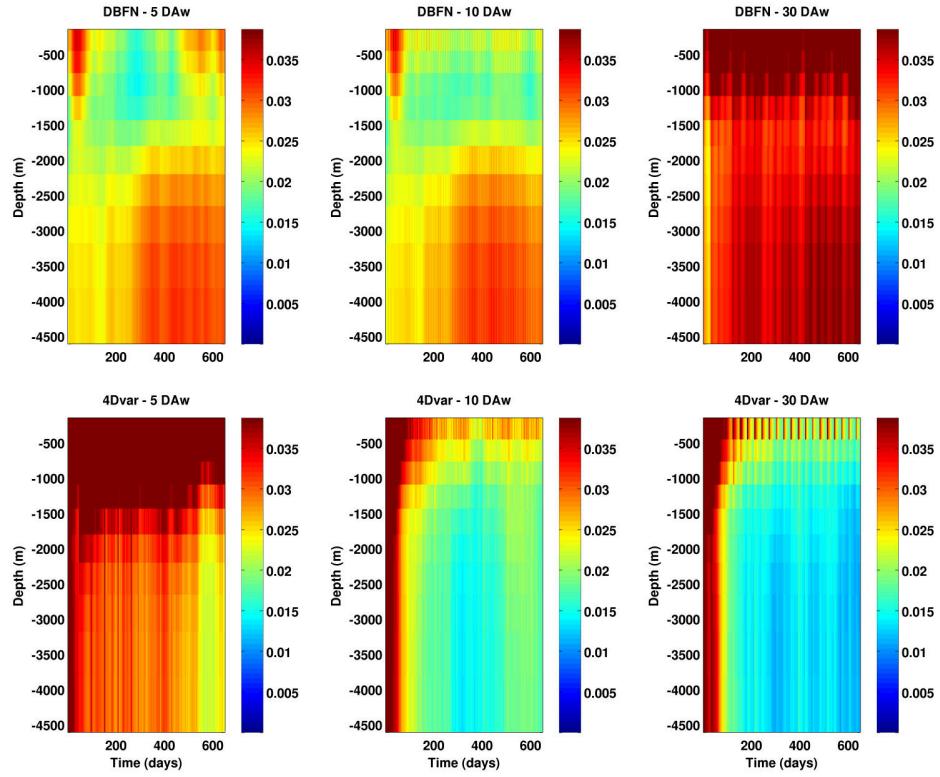


Fig. 11. Time evolution of vertical profiles of horizontally layer-wise averaged rms error of zonal velocities for the DBFN (top panels) and 4Dvar (bottom panels) experiments. Units are in (m/s).

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

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

566

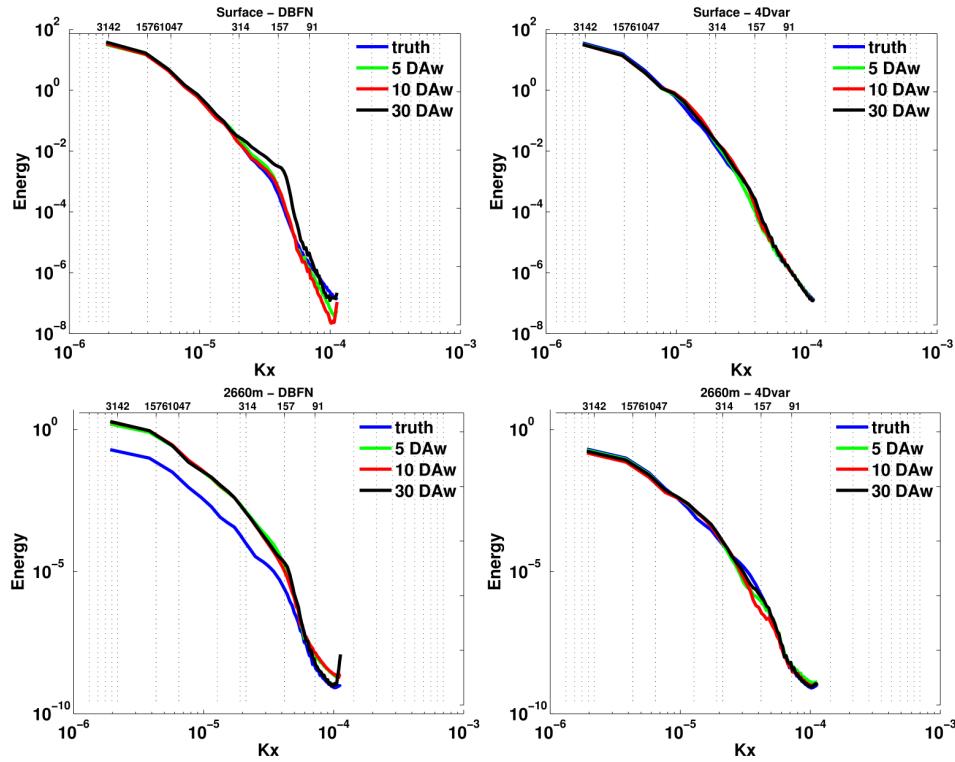


Fig. 12. Kinetic energy mean power spectra calculated using the first layer (top) and a layer at 2660m (bottom) and using the 650 days of the assimilation experiments using the DBFN (left) and the 4Dvar (right). Blue curves represent the “true” power spectra; Green curves represent the power spectra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

567 6.2.2 Observations density and first guess

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

576 Fig.13 shows the rms error for the SSH and zonal velocity. The results are quite similar to the
 577 results presented in Sect 6.1 with a lower rms error for all variables for both methods. Fig.14 shows
 578 the initial condition error for the zonal velocity produced by both methods for the satelite-like obser-

579 vations and the complete observations experiments. The figure reveals that while in some places the
 580 inclusion of more observations helps to reduce the error in other places it increases the error. This
 581 means that although much more information could be extracted from the new set of observations,
 582 which decreases the global rms errors, the solution remains dependent on the covariance structures
 583 contained on \mathbf{B} and $\hat{\mathbf{B}}^{PLS}$.

584

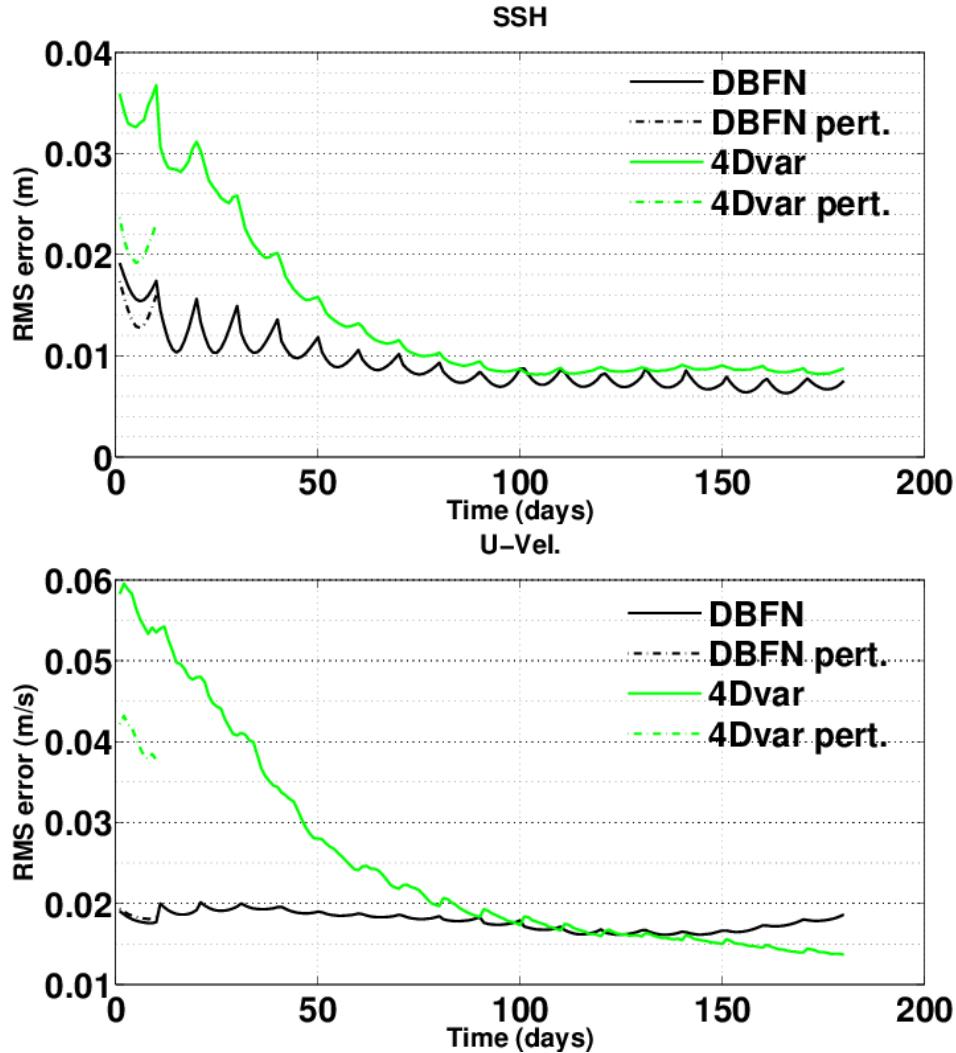


Fig. 13. RMS errors of SSH (top panel) and zonal velocity (bottom panel) from the DBFN and 4Dvar experiments with DAw of 10 days and assimilating complete daily fields of SSH. Dashed lines concern the results from the perturbed experiments.

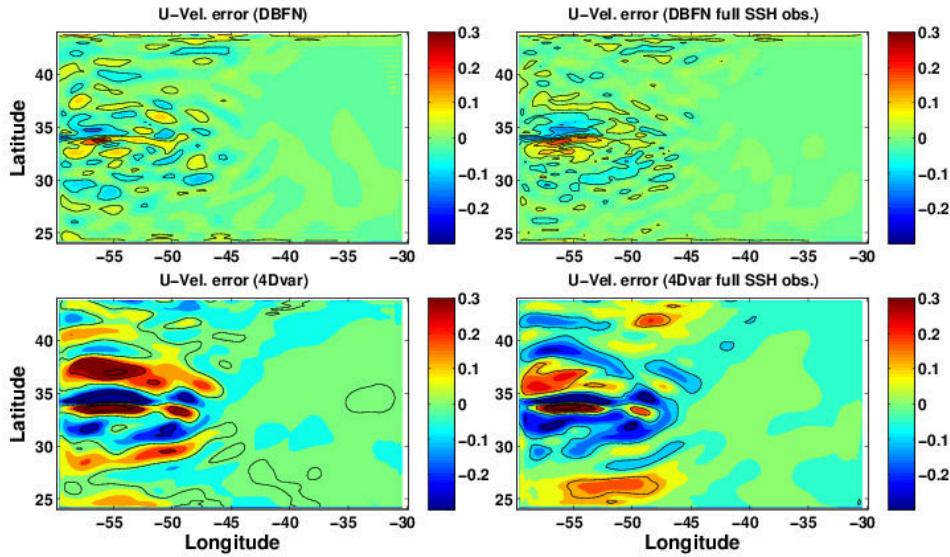


Fig. 14. Zonal velocity error (analysis - truth) for the first assimilation cycle from DBFN experiments (top panels) and 4Dvar experiments (bottom panels). Right panels show the results obtained by assimilating complete daily fields of SSH and the left panels the results from the experiments presented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

585 The second experiment is initialized with an initial condition that is different from the one used
 586 previously. We call this experiment as perturbed experiment. In this case, the objective is to analyze
 587 the sensitivity of the solution to the choice of the first guess. Thus, only one assimilation cycle is
 588 performed.

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

593 The DBFN produced smaller differences between the perturbed and non-perturbed experiences
 594 than the 4Dvar for the entire domain. A remarkable difference between the errors produced by the
 595 4Dvar and the DBFN is the error structure in the western boundary that is produced by the DBFN,
 596 which is positive northward $34^{\circ}N$ and negative southward $34^{\circ}N$. The presence of this structure is
 597 related to the fact that the DBFN analysis is the final condition produced by the backward model. The
 598 same pattern was also observed in the Fig. 3 that shows the backward error for the SSH variable.
 599 Since this region is a stable region, e.g. there are no meanders and vortices produced there, this
 600 suggests that the remaining errors produced by the DBFN project mostly onto the stable manifold
 601 as suggested by Auroux (2009). This should partially explain the other differences between the
 602 remaining errors produced by both methods as well as the better performance of the DBFN in the
 603 first assimilation cycle since the DBFN naturally corrects the forward unstable errors during the
 604 backward integration. The rms error of the identified trajectory for the perturbed experiment may be

605 seen in Fig. 13 as the green (4Dvar) and black (DBFN) dashed curves. The results clearly show that
 606 for the configured experiments the DBFN is much less sensitive to the first guess than the 4Dvar.

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

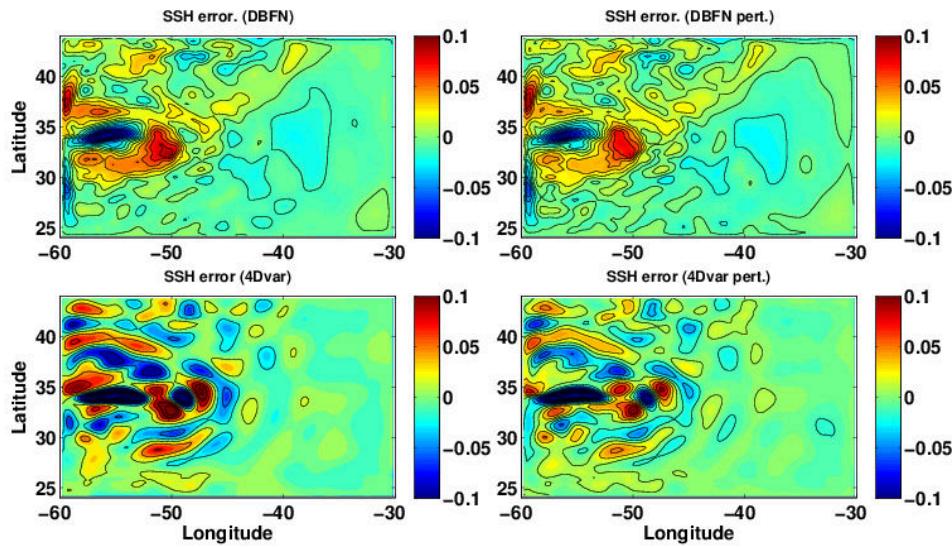


Fig. 15. SSH error (analysis - truth) from DBFN experiments (top panels) and 4Dvar experiments (bottom panels). Right panels show the results obtained from the perturbed experiment.

614 7 Conclusions and perspectives

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

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

625 Our results show that the DBFN can produce results comparable to 4Dvar using lower computa-

626 tional power. This is because DBFN demands less iterations to converge and because one iteration
627 of 4Dvar corresponds to one integration of the tangent linear model, one integration of the adjoint
628 model, which costs four times more than one standard model integration, plus the cost of minimizing
629 the cost function, while the DBFN costs twice the integration of the nonlinear model.

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

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

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

643 *Acknowledgements.* This work was supported by CNRS/INSU through the LEFE/MANU program. This work
644 was granted access to the HPC and visualization resources of "Centre de Calcul Interactif" hosted by "Université
645 Nice Sophia Antipolis ". Calculations were also performed at the IDRIS computational facility center. The
646 authors thanks Pierre-Antoine Bouttier for his help with the set-up of the 4Dvar algorithm.

647 **References**

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

650

651 Anderson, J. L.: A local least squares framework for ensemble filtering, Monthly Weather Review, 131, 634–
652 642, 2003.

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

655

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

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

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

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

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

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

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

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

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

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

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

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

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

683 Chen, X., Liu, C., ODriscoll, K., Mayer, B., Su, J., and Pohlmann, T.: On the nudging terms at open boundaries
684 in regional ocean models, Ocean Modelling, 66, 14 – 25, doi:10.1016/j.ocemod.2013.02.006, <http://www.sciencedirect.com/science/article/pii/S1463500313000401>, 2013.

685

686 Clifford, M., Horton, C., Schmitz, J., and Kantha, L. H.: An oceanographic nowcast/forecast system for the

687 Red Sea, Journal of Geophysical Research: Oceans, 102, 25 101–25 122, doi:10.1029/97JC01919, http:
688 //dx.doi.org/10.1029/97JC01919, 1997.

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

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

693 Donovan, A., Mirrahimi, M., and Rouchon, P.: Back and forth nudging for quantum state reconstruction, in: 4th
694 International Symposium on Communications, Control and Signal Processing, pp. 1–5, Limassol, Cyprus,
695 2010.

696 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo
697 methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, doi:
698 10.1029/94JC00572, http://dx.doi.org/10.1029/94JC00572, 1994.

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

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

705 Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local
706 ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112 – 126, doi:10.1016/j.physd.
707 2006.11.008, http://www.sciencedirect.com/science/article/pii/S0167278906004647, 2007.

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

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

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

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

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

722 Le Dimet, F. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological obser-
723 vations, Tellus, 38A, 97–110, 1986.

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

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

729 Leredde, Y., Devenon, J.-L., and Dekeyser, I.: Turbulent viscosity optimized by data assimilation, *Annales Geo-*
730 *physicae*, 17, 1463–1477, doi:10.1007/s00585-999-1463-9, <http://www.ann-geophys.net/17/1463/1999/>,
731 1999.

732 Levy, M. M., Klein, P., Triguier, A.-M., Iovino, D., Madec, G., Masson, S., and K. Takahashi, S.: Modifications
733 of gyre circulation by sub-mesoscale physics, *Ocean Modelling*, 34, 1 – 15, doi:10.1016/j.ocemod.2010.04.
734 001, <http://www.sciencedirect.com/science/article/pii/S1463500310000582>, 2010.

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

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

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

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

744 Luenberger, D. G.: Observers for Multivariable Systems., *IEEE Transactions on Automatic Control*, 11, 190–
745 197, doi:10.1109/TAC.1966.1098323, 1966.

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

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

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

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

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

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

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

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

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

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

767 observer, *Automatica*, 2010.

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

770 Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, *Mon. Wea. Rev.*, 132,
771 3019–3032, 2004.

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

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

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

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

780 Vidard, P. A., Le Dimet, F.-X., and Piacentini, A.: Determination of optimal nudging coefficients, *Tellus A*,
781 55, 1–15, doi:10.1034/j.1600-0870.2003.201317.x, <http://dx.doi.org/10.1034/j.1600-0870.2003.201317.x>,
782 2003.

783 Wang, K., Debernard, J., Sperrevik, A. K., Isachsen, E., and Lavergne, T.: A combined optimal interpolation
784 and nudging scheme to assimilate OSISAF sea-ice concentration into ROMS, *Annals of Glaciology*, 54,
785 8–12, 2013.

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

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

791 Willians, P. D.: A Proposed Modification to the Robert Asselin Time Filter., *Month. Weather Rev.*, 137, 2538–
792 2546, 2009.

793 Zou, X., Navon, I. M., and Le Dimet, F. X.: An Optimal Nudging Data Assimilation Scheme Using Param-
794 eter Estimation, *Quarterly Journal of the Royal Meteorological Society*, 118, 1163–1186, doi:10.1002/qj.
795 49711850808, <http://dx.doi.org/10.1002/qj.49711850808>, 1992.