Adjoint-Free 4D Variational Data Assimilation into Regional Models

M. Yaremchuk, P. Martin, G. Panteleev, C. Beattie and A. Koch

Abstract The ongoing trend towards parallelization in computer technologies propels ensemble methods toward the forefront of data assimilation studies in geophysics. Of particular interest are ensemble techniques which do not require the development of tangent linear numerical models and their adjoints for optimization. These "adjoint-free" methods detect effective search directions for optimization through direct perturbation of the numerical model across carefully chosen sets of states. Optimization proceeds by minimizing the cost function within the sequence of subspaces spanned by these perturbations. In this chapter, an adjoint-free variational technique (a4dVar) is described and demonstrated in an application estimating initial conditions of two numerical models: the Navy Coastal Ocean Model (NCOM), and the surface wave model (WAM). It is shown that a4dVar is capable of providing forecast skill similar to that of conventional 4dVar at comparable computational expense while being less susceptible to excitation of ageostrophic modes that are not supported by observations. Prospects of further development of the a4dVar methods are discussed.

M. Yaremchuk (✉) · P. Martin
Naval Research Laboratory, Stennis Space Center, Bay st louis, MS 39529, USA
e-mail: max.yaremchuk@nrlssc.navy.mil

G. Panteleev
International Arctic Research Center,
University of Alaska, Fairbanks, AK 99783, USA

C. Beattie
Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA

A. Koch
Department of Marine Science, University of Southern Mississippi,
Hattiesburg, MS 39529, USA

S.K. Park and L. Xu (eds.), *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III)*,
DOI 10.1007/978-3-319-43415-5_4

1 Introduction

As the speed of a single processor reached its physical limit of around 3GHz, the general trend in computer development in the last decade has moved from chips containing several cores to ones with tens or even tens of thousands of cores. In addition, multi-core chips mixed with simultaneous multithreading, memory-on-chip, and special-purpose heterogeneous cores promise further performance and efficiency gains in processing problems which can be efficiently split into parallel subtasks. In that respect, the maximum improvement that can be achieved in running atmospheric and oceanic models is limited by the number of grid points, n_ℓ, that can be attributed to a single core without incurring significant performance loss from inter-core communication.

These new capabilities and the increase in computational power they have produced have stimulated the development of *ensemble methods* for data assimilation. In contrast to adjoint-based 4d variational (4dVar) methods which run the numerical model and its adjoint in sequence in order to compute the cost function gradient, ensemble methods directly aggregate ensemble perturbations to acquire information on the cost function gradient and Hessian structure. In that respect, the ensemble-based 4dVar techniques offer significant parallel performance advantages, replacing the sequence of forward/adjoint model runs with up to M^2/n_ℓ parallel subtasks that involve only the forward model.

A related advantage that ensemble approaches offer is that they are *nonintrusive*, offering the opportunity to treat the numerical model as a black box and thus avoid the burdensome development and maintenance of tangent linear and adjoint codes required by 4dVar methods. Employing this property, Anderson and co-workers (Anderson et al. 2009; Hoteit et al. 2013) developed the Data Assimilation Research Testbed (DART) system on the basis of the widely used Ensemble Kalman filter (EnKF).

There has been recent, significant progress in extending EnKF techniques into the particle filtering framework (Hoteit et al. 2012) and in coupling EnKF techniques with both 3d- and 4d-variational methods (e.g., Županski 2005; Liu et al. 2008; Zhang et al. 2009). Of particular interest for the adjoint-free approach that we present here has been the development of the Maximum Likelihood Ensemble Filter (Županski 2005) based on the explicit computation of the square root of the Hessian matrix restricted to a subspace spanned by ensemble members.

The merging of ensemble approaches with variational techniques has developed along two lines: (a) improvement of the background error covariances (BECs) through the introduction of ensemble-based estimates and their hybrid generalizations (Clayton et al. 2013; Kuhl et al. 2013), and (b) searching for the optimal solution within the subspaces spanned by the leading error modes of the BECs. This second line of approach has been pursued by many authors in the last decade (Liu et al. 2008; Zhang et al. 2009; Zhang and Zhang 2012; Trevisan et al. 2010) and implicitly assumes that the BEC structure is well described by these (possibly localized) BEC modes. More recently, the performance of a family of adjoint-free

methods (4dEnVar) based on a formulation by Liu et al. (2008) has been compared with the standard 4dVar techniques in the framework of idealized experiments with the Lorenz-05 model (Fairbairn et al. 2014). These results show significantly better performance of 4dEnVar for moderate-length assimilation windows with low-density observations. Desroziers et al. (2014) demonstrated a close relationship between the 4dEnVar and 4dVar state space formulations and compared various implementations of 4dEnVar with 4dVar in an idealized setting.

The developments described above mostly deal with meteorological applications, where ensembles are supported by significantly higher data densities than are available in oceanographic applications. High data density allows one to obtain reasonably good estimates of BECs from the ensemble using truncated representation of the localization matrices and to efficiently compute the cost function gradient directly from ensemble perturbations (Liu et al. 2009; Tian and Xie 2012). A significant advantage of such an approach is the absence of the necessity to develop and maintain tangent linear and adjoint codes and the flexibility that results in adapting to various dynamical constraints.

In the ocean, ensemble-based BEC estimates tend to be less accurate, and one has to rely on *ad hoc* BEC representations (Mirouze and Weaver 2010; Yaremchuk and Sentchev 2012). Without reliable correlation information, the development of an efficient adjoint-free assimilation method also becomes more problematic as one must select a small number of reliable perturbations with more care. Early attempts to develop practical a4dVar algorithms in oceanography were limited to predetermined low-dimensional subspaces spanned either by the reduced-order approximations of the model Green's functions (Stammer and Wunsch 1996; Menemenlis and Wunsch 1997), or by the dominant principal component vectors (EOFs) associated with the model statistics (Qui et al. 2007; Hoteit 2008). In fact, the 4dEnVar technique pursues a similar, but more general approach, parameterizing the search subspace by Schur products of the ensemble members with the eigenvectors of the reduced-order representation of the localization matrix.

In this chapter, we give an overview of recent developments in adjoint-free methods of data assimilation using both ensemble-generated and *ad hoc* BEC models, and illustrate the basic principles of the latter approach using an idealized optimization problem constrained by linear dynamics. We describe a particular approach to adjoint-free 4dVar, referred to here as a4dVar (cf., Yaremchuk et al. 2009), which we then apply in Sect. 3 to the assimilation of hydrographic surveys and velocity observations collected in the Adriatic Sea in 2006. Assimilation is constrained by the state-of-the-art Navy Coastal Ocean Model (NCOM) and a4dVar results are compared with those obtained by means of the traditional 4dVar technique. In Sect. 4 the a4dVar method is tested with simulated data constrained by a spectral surface wave model and the forecast skill of the optimized solution is compared with one delivered by an operational method based on sequential assimilation of significant wave height. Section 5 completes the chapter with a summary and discussion of the prospects for adjoint-free methods in general, and a4dVar in particular.

2 Variational Data Assimilation

2.1 Adjoint Methods

Consider the 4dVar method as solving as the following linear discrete least-squares problem constrained by model dynamics in a small vicinity of the model's background trajectory $\mathbf{x}_b^n$:

$$J = \frac{1}{2}\left[\mathbf{x}^{0\mathsf{T}}\mathbf{B}^{-1}\mathbf{x}^0 + \sum_{n=0}^{N}(\mathbf{H}_n\mathbf{x}^n - \mathbf{d}^n)^{\mathsf{T}}\mathbf{R}_n^{-1}(\mathbf{H}_n\mathbf{x}^n - \mathbf{d}^n)\right] \to \min_{\mathbf{x}^0}. \tag{1}$$

Here n enumerates observation times, $\mathbf{B}$ is the error covariance matrix of $\mathbf{x}_b^0$ which describes the (Gaussian) error statistics of the model state at $n = 0$, $\mathbf{H}_n$ are the model-data projection operators, $\mathbf{d}^n$ are the misfits $\mathbf{y}_n^o - \mathbf{H}_n\mathbf{x}_b^n$ between observations $\mathbf{y}_n^o$ and the corresponding background model values, $\mathbf{R}_n$ are the observation error covariances, and $^{\mathsf{T}}$ denotes transposition. We will denote the dimension of the discretized model state vector $\mathbf{x}$ by M and the total number of observations by M_d.

The optimal correction vector $\mathbf{x}^n$ is governed by the recursive relationship $\mathbf{x}^n = \mathbf{M}_n\mathbf{x}^{n-1}$, where $\mathbf{M}_n$ is the dynamical operator of the model linearized in the vicinity of the background trajectory, $\mathbf{x}_b^n$, across the time interval (t_{n-1}, t_n), so that

$$\mathbf{x}^n = \mathbf{M}_n\mathbf{M}_{n-1}\ldots\mathbf{M}_2\mathbf{M}_1\mathbf{x}^0. \tag{2}$$

To avoid the ensuing clutter of symbols, we introduce new notation: $\mathbf{c} \equiv \mathbf{x}^0$ for the control vector, $\mathbf{M}^n \equiv \mathbf{M}_n\ldots\mathbf{M}_2\mathbf{M}_1$ for the aggregated n-step propagator, $\overline{\mathbf{H}}_n \equiv \mathbf{R}_n^{-1/2}\mathbf{H}_n$, $\overline{\mathbf{d}}^n \equiv \mathbf{R}_n^{-1/2}\mathbf{d}^n$. We then drop the over-bars and so, taking (2) into account, the minimization problem (1) can be rewritten in terms of the optimal correction, $\mathbf{c}$, to the initial state:

$$J = \frac{1}{2}\left[\mathbf{c}^{\mathsf{T}}\mathbf{B}^{-1}\mathbf{c} + \sum_{n=0}^{N}(\mathbf{H}_n\mathbf{M}^n\mathbf{c} - \mathbf{d}^n)^{\mathsf{T}}(\mathbf{H}_n\mathbf{M}^n\mathbf{c} - \mathbf{d}^n)\right] \to \min_{\mathbf{c}}. \tag{3}$$

A 4dVar data assimilation method finds the minimum of J by solving the normal equations, expressed as:

$$\nabla_{\mathbf{c}} J = \mathbf{B}^{-1}\mathbf{c} + \sum_n \mathbf{M}^{n\mathsf{T}}\mathbf{H}_n^{\mathsf{T}}(\mathbf{H}_n\mathbf{M}^n\mathbf{c} - \mathbf{d}^n) = 0, \tag{4}$$

To further simplify discussion, introduce the following notation for the Hessian matrix, $\tilde{\mathbf{H}}$, and the right-hand side, $\mathbf{b}$,

$$\tilde{\mathbf{H}} = \mathbf{B}^{-1} + \sum_n \mathbf{M}^{n\mathsf{T}} \mathbf{H}_n^{\mathsf{T}} \mathbf{H}_n \mathbf{M}^n; \quad \mathbf{b} = \sum_n \mathbf{M}^{n\mathsf{T}} \mathbf{H}_n^{\mathsf{T}} \mathbf{d}^n, \tag{5}$$

allowing us to rewrite the normal equations as $\tilde{\mathbf{H}}\mathbf{c} = \mathbf{b}$.

There are two major approaches to 4dVar data assimilation. The first one, the *state space approach*, iteratively solves (4) through a conjugate gradient descent or related algorithm, which on every iteration computes the gradient and then estimates an effective descent direction using information on the Hessian accumulated in previous iterations. This method is widely used in a number of community OGCMs (NEMO, ROMS), and in operational meteorology (ECMWF).

As may be seen from (4), this process must involve the application of both the model evolution operator, $\mathbf{M}^n$, and its transpose, $\mathbf{M}^{n\mathsf{T}}$ (the "adjoint model"). The numerical procedure of calculating the gradient involves two major steps:

(1) Sequential calculation of $\mathbf{x}_i^n = \mathbf{M}^n \mathbf{c}_i$ (forward run of the tangent linear model), supplemented additionally by the calculation of the quantities $\mathbf{q}_i^n = \mathbf{D}_n \mathbf{x}_i^n - \mathbf{H}_n^{\mathsf{T}} \mathbf{d}^n$.

(2) Accumulation of the products $\mathbf{M}^{n\mathsf{T}} \mathbf{q}_i^n$ conveniently performed in the reverse-time order because $\mathbf{M}^{n\mathsf{T}} = (\mathbf{M}_n \ldots \mathbf{M}_1)^{\mathsf{T}} = \mathbf{M}_1^{\mathsf{T}} \ldots \mathbf{M}_n^{\mathsf{T}}$ (backward-in-time integration of the adjoint model).

The sequential nature of this algorithm generally will limit parallel scalability.

A second approach to 4dVar data assimilation is the *observation space approach*—so called because the solution process is mapped into the space of observations instead of remaining solely in the state space. The framework for this may be developed by using the Sherman-Morrison-Woodbury formula to transform the Hessian inverse from having action defined directly in the state space to equivalent action defined in the (generally lower dimensional) observation space. This transformation tactic is closely related to "optimal interpolation" as it seeks the optimal solution in the form of a linear function of model-data misfits, leading also to a family of methods called "representer methods" (Bennett 2002; Rosmond and Xu 2006). In most geophysical applications there will be significant benefit in searching for a solution in the M_d-dimensional observation space as opposed to the much larger M-dimensional state space. The observation space method also has a certain advantage over the state space approach with respect to parallel computing efficiency, since the computation of M_d representers can be performed independently and in parallel.

The robustness of final estimators may be improved if one separates the aggregate background error covariance term into a component, $\mathbf{B}_0$, associated with the uncertainty of the initial state, and components, $\mathbf{B}_n$, associated with the uncertainties of the model equation and forcing. In effect, one replaces $\mathbf{c}^{\mathsf{T}} \mathbf{B}^{-1} \mathbf{c}$ in the cost function (3) with $\mathbf{c}^{\mathsf{T}} \mathbf{B}_0^{-1} \mathbf{c} + \sum_{n \geq 1} \mathbf{e}^{n\mathsf{T}} \mathbf{B}_n^{-1} \mathbf{e}^n$ which now involves model errors, $\mathbf{e}^n = \mathbf{x}^n - \mathbf{M}_{n-1} \mathbf{x}^{n-1}$. The normal equations in this case are more complicated than

(4), but some numerical advantages accrue in approaching the resulting computational task using the representer method (Bennett 2002; Rosmond and Xu 2006). Minimization of the (non-linear) cost function in the observation space involves multiple convolutions with the (generally nonsparse) matrices $\mathbf{B}_n$, making the method sometimes more computationally expensive than the state space approach. This method has been implemented in ROMS (Moore et al. 2011) as an optional feature, and in the Naval Research Laboratory for both atmospheric (Xu and Rosmond 2004; Xu et al. 2005) and oceanic (Ngodock and Carrier 2014) data assimilation systems.

2.2 *Adjoint-Free Methods*

As the name suggests, adjoint-free methods perform minimization of the cost function without using linearized models and their adjoints. This is achieved by the direct assessment of cost function sensitivity through an ensemble of parallel model runs using perturbed control parameters. Assuming that control parameter perturbations capture the dominant modes of the background error statistics, the ensemble of model trajectories that is produced can be used to estimate the dynamically consistent evolution of the background error covariance, which is implicitly performed by the 4dVar algorithm during optimization.

Currently, the most developed adjoint-free technique is the $\mathbf{B}$-preconditioned state space approach proposed by Liu et al. (2008, 2009), referred to as 4dEnVar in literature. The major idea is to seek the 4dVar solution in the subspace spanned by the model perturbations. This makes the method equivalent to the observation space 4dVar with the only difference that the search is executed in $\mathbb{R}^{km}$, where m is the ensemble size and k is the number of eigenmodes in the covariance localization matrix used to diversify search directions (Hamill et al. 2001; Liu et al. 2009). Currently, the method has been successfully tested with real data (Liu et al. 2013) and in Meteo France/UK Met Office (Fairbairn et al. 2014; Desroziers et al. 2014) in a more theoretical context.

Although 4dEnVar has shown promise, the method has some deficiencies which may hinder its use, especially in oceanographic practice, where observations are not as plentiful as in atmospheric practice and, as a consequence, ensembles may be insufficiently accurate in approximating the background error statistics. One may also note that the 4dEnVar minimization process is still based on the sequential computation of gradients used in the course of building the optimal solution; this could lead to a performance bottleneck in massively parallel computing environments.

We discuss another method of adjoint-free minimization based on projecting $\tilde{\mathbf{H}}$ onto the subspace spanned by ensemble perturbations. The approach was first utilized in the Maximum Likelihood Ensemble Filter (Županski 2005) and later extended to an adjoint-free 4dVar variational algorithm that we will refer to as a4dVar (Yaremchuk et al. 2009; Panteleev et al. 2015). The technique involves direct minimization of the cost function in a sequence of $\tilde{\mathbf{H}}$-orthogonal subspaces and requires an efficient algorithm for computing the action of $\mathbf{B}^{-1/2}$ on a vector which nicely fits

the approach in heuristic BEC modeling using polynomials of the diffusion operator (Yaremchuk et al. 2013; Yaremchuk and Sentchev 2012). Although the a4dVar formulation guarantees its convergence in M/m iterations, practical feasibility requires obtaining a reasonable degree of accuracy in solving the normal equation within several dozen iterations. This is achieved by restricting the basis vectors of the search subspaces to be smooth, implicitly assuming that the leading eigenvectors of $\tilde{\mathbf{H}}^{-1}$ have this property and that the rhs $\mathbf{b}$ of the normal equation will have a sizable projection onto this "smooth manifold".

To illustrate these ideas, consider a simple problem of retrieving the initial field of tracer concentration $c(\boldsymbol{x}, 0)$ from observations at some distant time T. The tracer evolution is governed by

$$\partial_t c + \boldsymbol{u}\nabla c - \mu\Delta c = f(\boldsymbol{x}, t) \tag{6}$$

in a closed rectangular 49×91 domain Ω (Fig. 1) with the boundary condition $\eta(\partial\Omega, t) = 0$. Equation (6) is discretized on a regular grid using simple first-order explicit time-stepping, upwind advection, and a standard 5-point stencil for the Laplacian with unit steps in temporal and spatial directions. The velocity $\boldsymbol{u} = (u, v)$ at any space-time location is defined by $u = -0.2 + 0.01\nu; \quad v = -0.1 + 0.01\eta$, where η is the white noise on unit interval. The forcing f is generated by setting $f(\boldsymbol{x}, t) = 0.001\eta$ in every point of the space-time grid. The coefficient μ is set to 10^{-5}, so that diffusion is largely determined by the numerics.

The simulated data experiment is set as follows: Given the initial tracer distribution $\hat{c} = c(\boldsymbol{x}, 0) = \exp[-(\boldsymbol{x} - \boldsymbol{x}_0)^2/9]$ with $\boldsymbol{x}_0 = (70, 35)$ (bell-shaped disturbance in Fig. 1a), the model is integrated for $T = 200$ time steps to obtain the final distribution $c(\boldsymbol{x}, T)$ shown by contours in the same panel. Notice that the initial disturbance almost completely dispersed and migrated to $\boldsymbol{x}_T \sim (25, 15)$. After that, $c(\boldsymbol{x}, T)$ is sampled at 200 points shown in Fig. 1a, and the numbers obtained are used to reconstruct $\hat{c}$ by minimizing the cost function (3) under the dynamical constraint (6) with an inverse background error covariance defined by

$$\mathbf{B}^{-1} = \left[\mathbf{I} - \frac{a^2}{2}\Delta\right]^2 \tag{7}$$

where $\mathbf{I}$ is the identity operator in state space and $a = 1.5$ is the decorrelation scale. With the definition (7) at hand, it is easy to compute the action $\tilde{\mathbf{H}}^{1/2}$ on a control vector and perform $\tilde{\mathbf{H}}$-orthogonalization (see Appendix).

For the purpose of comparison, the cost function is minimized using the state-space 4dVar technique and two versions of a4dVar, which differ in the method of building the search directions (SDs). The number m of SDs (ensemble size) in both a4dVar versions is set to 10. The first version specified SDs as a sequence of tens of eigenvectors of $\mathbf{B}$ in descending order of eigenvalue magnitude. To specify search directions for the second a4dVar method, 200 observations were split into $m = 10$ equal groups so an observation operator $\mathbf{H}_j$ for the jth search direction is sampling

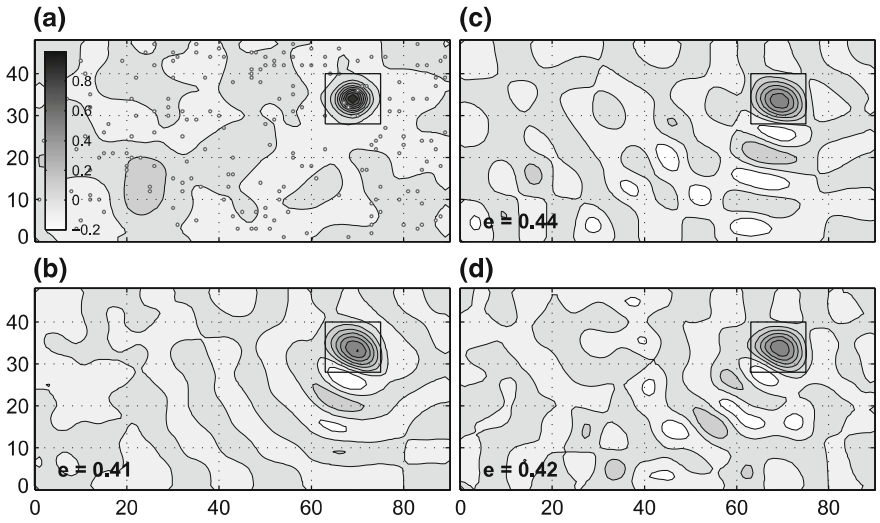

Fig. 1 Reconstruction of the initial condition of the tracer field by 4dVar (**b**) and a4dVar (**c**, **d**) techniques. Composite map of the tracer field evolution being reconstructed is shown in *panel* a with the initial position of the reconstructed feature (Gaussian eddy at $\boldsymbol{x} = (70, 35)$) superimposed on the tracer field (contours) at the observation time ($t = 200$). *Circles* denote observation points. The errors in approximation of the true perturbation at $t = 0$ are shown in the *left corner*

a group of 20 distinct locations among those shown in Fig. 1a. SDs $\mathbf{s}_j$ on the ith iteration were defined by

$$\mathbf{s}_j^i = (\mathbf{B}^{-1} + \mathbf{H}_j^{\mathrm{T}}\mathbf{H}_j)^{-1}\mathbf{q}_j^i, \qquad j = 1, ..., 10 \tag{8}$$

Optimal approximations $\tilde{c}$ to $\hat{c}$ obtained by 4dVar and a4dVar techniques are shown in Fig. 1b and Fig. 1c, d respectively). The quality of reconstruction was assessed by the parameter

$$e = \sqrt{\langle(\tilde{c} - \hat{c})^2\rangle / \langle\hat{c}^2\rangle} \tag{9}$$

where angular brackets denote averaging over rectangles enveloping the reconstructed perturbation in Fig. 1. Comparison of Fig. 1b, c, d suggests that the a4dVar method is capable of providing a solution of the same quality with 4dVar, and that the general a4dVar strategy of minimizing J using a sequence of smooth $\tilde{\mathbf{H}}$-orthogonal SDs may work well with various methods of generating the ensemble members.

In terms of computational expense, the 4dVar method provided approximately five times faster reduction of the cost function (Fig. 2) due to high efficiency of the adjoint model. In this simple case, an adjoint model run required the same amount of time as the direct model run. In real applications, the tangent linear and adjoint codes are several times more expensive to run and the a4dVar techniques may prove to be more competitive, as shown in Sect. 3.

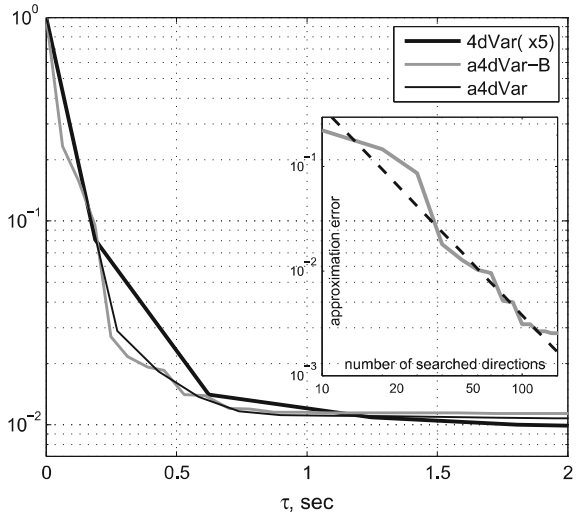

Fig. 2 Reduction of the cost function against CPU time for 4dVar and a4dVar techniques. The 4dVar CPU time is multiplied by five to mimic larger CPU requirements of the state-of-the-art adjoint models. Inset: Convergence of the a4dVar-B solution (Fig. 1c) to the exact solution. *Dashed line* shows the convergence rate given by (12)

The well-known structure of $\mathbf{B}$ provides an opportunity to assess the convergence rate of the ad4Var solution exposed in Fig. 1c. Assume that after k a4dVar iterations $m_s = km$ $\tilde{\mathbf{H}}$-orthogonal directions have been already searched and the kth approximation $\hat{c}_k$ to the optimal solution $\hat{c} = \tilde{\mathbf{H}}^{-1}\mathbf{b}$ have been found. Without loss of generality, the eigenvectors ϕ_i of $\mathbf{B}$ could be normalized to satisfy $\phi_i \mathbf{B}^{-1}\phi_i = 1$, so that their (Euclidean) norm is equal to the associated eigenvalue σ_i. The magnitude e_{m_s} of the approximation error $\mathbf{e}_k = \hat{c} - \hat{c}_k$ with respect to the norm induced by the inverse covariance can be assessed by projecting $\hat{c}$ on the *unexplored* directions:

$$e_{m_s} = \mathbf{e}_k^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{e}_k \le \sum_{l>m_s} |\hat{c}^{\mathrm{T}}\mathbf{B}^{-1}\phi_l|^2 \tag{10}$$

Furthermore, since the optimal solution $\hat{c} = \tilde{\mathbf{H}}^{-1}\mathbf{b}$ allows representation in the (dual) form $\hat{c} = \mathbf{B}\rho$ (ρ is the optimal linear combination of the representers), the upper bound of the terms under summation in (10) can be assessed by

$$|\hat{c}^{\mathrm{T}}\mathbf{B}^{-1}\phi_l| = |\rho^{\mathrm{T}}\phi_l| \le \sigma_l(\rho^{\mathrm{T}}\mathbf{B}^{-1}\rho)^{1/2} \tag{11}$$

Plugging (11) into (10) yields the following upper bound on the error magnitude:

$$e_{m_s} \le \rho^{\mathrm{T}}\mathbf{B}^{-1}\rho \sum_{l>m_s} \sigma_l \sim O(m_s^{-2}) \tag{12}$$

This estimate remains intact if we assess e_{m_s} with respect to the norm induced by the Hessian matrix. In the latter case, the right-hand side of (12) will be additionally multiplied by a scaling factor $||\tilde{\mathbf{H}}||/||\mathbf{B}^{-1}|| > 1$.

Dependence of the distance between the 4dVar solution (Fig. 1b) and the consecutive approximations to the a4dVar solution (Fig. 1d) shown in the inset to Fig. 2, confirms the above estimate.

Similar experiments with a low-dimensional ($M = 1,922$) non-linear quasigeostrophic model were performed by Yaremchuk et al. (2009) who documented compatible performance of the 4dVar and a4dVar methods in the non-linear regime and certain advantages of the a4dVar approach in the cases of sparse and/or noisy observations. In the next sections we present the results of applying a4dVar to real and simulated data constrained by state-of-the-art numerical models.

3 a4dVar and 4dVar Assimilation of Real Data in the Adriatic Sea

3.1 Model and Data

The NCOM is a free-surface primitive-equation hydrostatic ocean model with σ coordinates in the upper layers and, optionally, fixed depths below a user-specified distance from the surface. Algorithms that comprise a NCOM computational kernel are described in Martin (2000); Barron et al. (2006). The model was configured at 3 km resolution on an 85 × 294 horizontal grid (Fig. 3) with 32 levels in the vertical. The top 22 σ levels follow the bathymetry, stretching from the surface to a fixed depth of 291 m, and 10 fixed-depth levels are used below 291 m. Initial and open boundary conditions for the sea surface height ζ, temperature T, salinity S, and horizontal velocities u, v were provided from the regional NCOM simulation (Martin et al. 2009). The model was forced by the river runoff and atmospheric fluxes derived from the regional atmospheric model with 8 km horizontal resolution. In the described assimilation experiments, initial conditions were used as control variables,

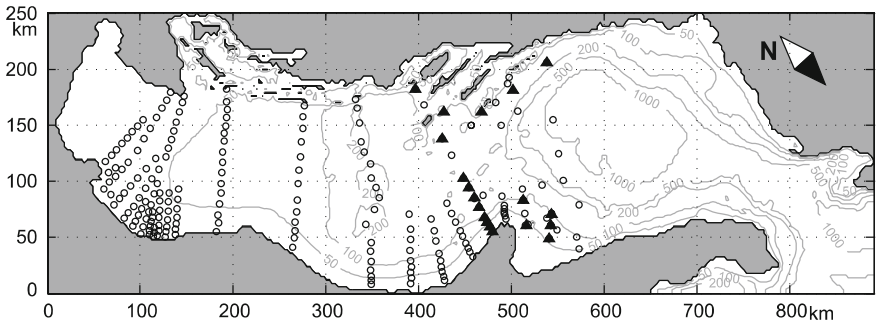

Fig. 3 Model domain with CTD stations (*circles*) and moorings (*triangles*) of the DART experiment. *Gray* contours (m) show *bottom* topography

i.e., the vector $\mathbf{c}$ comprised all the grid point values of ζ, T, S, u, v at $n = 0$. With the given 3-dimensional grid and bathymetry, the inverse problem has $M = 1{,}493{,}570$ unknowns.

The first guess (background) values of $\mathbf{c}$ were taken from the NCOM simulation described by Martin et al. (2009) and then adjusted to suppress temperature and salinity biases during the assimilation time interval (0.00 UTC on 08/14 to 0.00 UTC on 08/29/2006). After the adjustment, the horizontal-and-time average misfits between the background solution and *TS* observations did not exceed 0.02 °C and 0.005 psu, respectively.

Assimilated data were acquired in the course of the field experiment Dynamics of the Adriatic in Real Time (DART) (Martin et al. 2009; Burrage et al. 2009). In the present study, CTD and ADCP observations from August 14 to August 29, 2006 are used (Fig. 3). Temperature T and salinity S were measured at 219 CTD stations occupied in the northern and central parts of the basin. The total number of TS observations used in the assimilation is 9,650. Current velocities u, v were measured by 19 moored ADCPs in the depth range from 15 to 150 m at locations shown by triangles in Fig. 3. All the velocity data were detided and averaged over 29 twelve-hour intervals centered at the assimilation times t_n of 0 and 12 UTC. With the total number of the observed velocities 13,856 the dimension of the observation space was M_d =23,506.

3.2 Assimilation Parameters

In the course of the experiments the parameters of the tested 4dVar and a4dVar systems were kept as close as possible to each other. However, due to the different formulations (observation space for NCOM 4dVar and state space for a4dVar), certain discrepancies remained in the shape of the background error covariance $\mathbf{B}$. In both algorithms $\mathbf{B}$ is given by the product $\mathbf{VCV}$ where $\mathbf{V}$ is the diagonal matrix of the background error rms variances and $\mathbf{C}$ is the respective correlation matrix.

In the 4dVar algorithm, the action of $\mathbf{C}$ on a state vector is represented by the operator

$$\mathbf{C} \simeq \exp(\frac{1}{2}b^2\Delta), \tag{13}$$

which is implemented numerically by integrating the heat transfer equation (e.g., Weaver and Courtier 2001) with the decorrelation length scale $b = 9$ km. Since the matrix $\mathbf{C}$ is rank-deficient, the second-order polynomial approximation to the exponent in (13) was used to define $\mathbf{C}^{-1}$ in the a4dVar algorithm. Parameter a was set to $\sqrt{8/\pi}b$ to preserve the value of the integral decorrelation scale specified in 4dVar (e.g., Yaremchuk and Smith 2011). The rest of the assimilation parameters were identical for both the 4dVar and a4dVar assimilation systems.

The tested a4dVar method is based on Eq. 8 and outlined as follows:

0. Specify the dimension m_s of the search subspaces, their number k to be kept in memory for $\tilde{\mathbf{H}}$-orthogonalization, the maximum number of iterations I, the perturbation magnitude ε and the background model trajectory $\mathbf{x}_b^n$. Set the iteration number i to zero, $\mathbf{c}_0 = 0$, and compute $\mathbf{d}^n$.

1. Compute $\mathbf{x}_i^n, J_i, \mathbf{Y}_i = \tilde{\mathbf{H}}^{1/2}\mathbf{c}_i$ and the search directions $\mathbf{s}_n^i, n = 0, \ldots, N$ (Eq. 8).

2. Extract the m_s leading EOFs $\mathbf{p}_i^m, m = 1, \ldots, m_s$ of the search directions to form the basis in the search subspace.

3. Perturb the initial conditions $\mathbf{c}_i \to \mathbf{c}_i + \varepsilon\mathbf{p}_i^m$ and run (in parallel) the ensemble of m_s perturbed models, computing the respective perturbed values of J_i^m and $\mathbf{Y}_i^m$.

4. $\tilde{\mathbf{H}}$-orthogonalize the search basis $\{\mathbf{p}_i^m\}$ with respect to at most k basis vectors obtained on the previous iterations and compute optimal corrections $\delta\mathbf{c}_i$ (see Appendix 1).

5. Set $\mathbf{c}_{i+1} = \mathbf{c}_i + \delta\mathbf{c}_i$.

6. If $i = I$ exit. Otherwise set $i \leftarrow i + 1$, then go to 1.

The stopping criteria for the iterative processes were selected as follows: For the 4dVar system the solution of the system for the representer coefficients was terminated after $n_t = 7$ iterations, when the accuracy of the conjugate gradient (CG) solver was, as a rule, better than 10^{-3}. With the value of $n_t = 7$, 8–10 outer loops were executed before the values of J started to increase. For the a4dVar system, the minimization was terminated when the total CPU time reached the value used by the respective 4dVar experiment. The number of ensemble members was kept constant at $m_s = 9$ through all the experiments.

3.3 Comparison with 4dVar

In the reported experiments we varied the length of the assimilation window from short (4 days, $N = 9$) to moderate (8 days, $N = 17$) and long (14 days, $N = 29$) duration. Performance of the assimilation algorithms was evaluated in three categories: the forecast skill at the end of the assimilation window (for $N = 9, 17$), the rate of convergence, and by qualitative inspection of the optimal model trajectories.

3.3.1 Convergence Rates and Computational Expense

To assess the rates of convergence, one has to have an ability to compare the reduction of the cost function with iterations, which is not straightforward for two reasons.

First, in the 4dVar algorithm considered here, the regularization term of the cost function can be evaluated only within the range of the correlation matrix defined by (13). To avoid the burden of restricting the a4dVar correlation matrix to the range of $\mathbf{C}$, we compared only the observational parts of the 4dVar and a4dVar cost functions (second term in Eq. (3)).

Second, the number of iterations required for convergence cannot be considered as an objective criterion because 4dVarV and a4dVar iterations are different in nature. Due to the non-linearity of the problem, an iteration (either 4dVar or a4dVar) performs minimization in the vicinity of the current (suboptimal) state, but 4dVar does that in the range of $\mathbf{B}$, whereas a4dVar minimizes in the subspace of a much smaller dimension spanned by $\mathbf{p}^m$. For that reason, iterations require quite different computational resources and should be compared in terms of CPU time. Figure 4 shows such a comparison by rescaling the horizontal axis with the total CPU time τ_a required by one a4dVar iteration. The value of τ_a was 11 times larger than the CPU time τ_m of a direct NCOM model run for a given experiment, i.e. $\tau_a \simeq 11\tau_m$. The major contribution to τ_a is given by the ensemble run ($9\tau_m$, p.3 in the layout of Sect. 3.2), while the master NCOM run (p.1) and operations listed in pp.2 and 4 require τ_m and $0.8\tau_m$, respectively. Overall, convergence was achieved at an expense of 60–70 iterations (650–800 NCOM runs).

As may be seen in Fig. 4, a single 4dVar iteration was approximately equivalent to 6–7 a4dVar iterations, or 70–80 direct model runs. This computational expense arises because sequential execution of the adjoint and tangent linear codes (inner loops of the CG solver) required around $11\tau_m$, whereas one 4dVar outer loop included seven inner loops to solve the system of linear equations for the representer coefficients.

Figure 4 shows that, in general, the tested a4dVar method is computationally comparable to the observation space 4dVar. Although the total CPU time required for reduction of J by the factor of 0.4 (attained after the first outer loop of the 4dVar) appears to be similar for the 4dVar and a4dVar methods, the a4dVar minimization noticeably slows down at subsequent iterations, especially for longer assimilation windows ($\tau = 8, 14$ days).

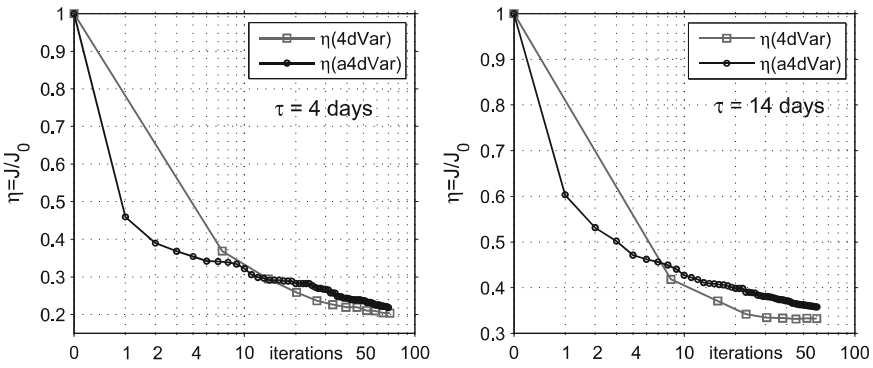

Fig. 4 Relative reduction η of the cost function with iterations (marked by *circles*) for different assimilation periods. The *horizontal axis* is scaled by the CPU time required for the a4dVar iteration. *Squares* label the 4dVar outer loops

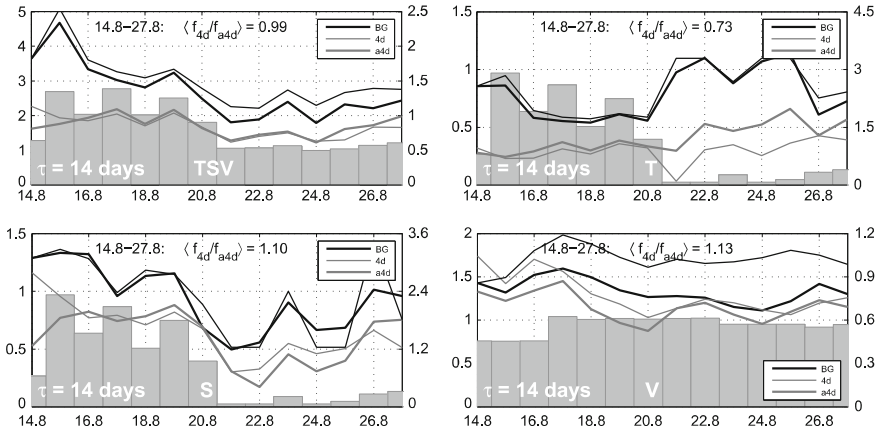

Fig. 5 Evolution of the root-mean-square model-data misfits f_q characterizing the background (BG, *thick lines*), 4dVar-optimized (4d, *thin gray lines*) and a4dVar-optimized (a4d) solutions. *Thin black line* shows the misfit with the background fields at $t = 0$ (persistence). The values of f_q are shown on the *right axis* of each panel. The *left axis* quantifies the number of the data points for each day in thousands (shown by *gray shaded rectangles*). The ratio of the mean values of f_q averaged over the assimilation window for the 4dVar and a4dVar methods is given

Figure 5 demonstrates the time evolution of the quantities

$$f_q^n = \langle \left[(\mathbf{H}_n \mathbf{x}_q^n - \mathbf{d}_q^n)^{\mathrm{T}} (\mathbf{H}_n \mathbf{x}_q^n - \mathbf{d}_q^n)/n_q \right]^{1/2} \rangle \tag{14}$$

characterizing the daily averaged $\langle\rangle$ model-data misfits of the various state vector components before (black lines) and after (gray lines) optimization with a 14-day assimilation window (i.e., using all the available data). The subscript q takes the values of the labels in the mid-bottom parts of Fig. 5 which indicate the observed variables (temperature, salinity and velocity vector) for which the statistics f_q were computed, whereas n_q stands for the total number of respective observations taken at a given day. The upper left panel in Fig. 5 shows a remarkable similarity in the time evolution of the combined model-data misfit for the 4dVar- and a4dVar-optimized NCOM states. The a4dVar algorithm has, however, a noticeable tendency to provide a better fit at the beginning of the assimilation window, clearly visible in the lower panels for f_S and f_v. This can be explained by the above mentioned property of a4dVar to better retrieve optimal states at shorter integration times.

When separated into different components, behavior of f_T^n, f_S^n, and f_v^n reveals more differences. In particular, the 4dVar method provides a much better fit to the temperature data after August 20 (in the second half of the assimilation window), but appears to be 10–13% worse than a4dVar with respect to salinity and velocity.

A large contribution to a better salinity fit is given by the first two days of the a4dVar model trajectory (lower left panel in Fig. 5). However, certain gains relative to 4dVar are also observed at the end of the assimilation, which is quite opposite to the difference in the values of f_T.

Compared to f_T and f_S, the overall improvement of the model-data misfit is the smallest for velocity (lower right panel in Fig. 5), which was characterized by the observation errors of $\mathbf{R}^{1/2} \sim 7$–10 cm/s in the cost function. Several assimilation runs were made with significantly smaller (3–5 cm/s) errors, but they were found to be inconsistent with a posteriori statistics of the model-data misfits as the optimal cost function values in these cases were much larger than those obtained in the reported experiments. The a4dVar-optimized value of f_v is persistently smaller (as compared to 4dVar) during the entire assimilation period providing the 13 % better value in the 14-day average. This advantage could be partly attributed to the fact that the a4dVar search directions are derived from the most persistent patterns of the model-data misfits and therefore tend to be closer to the slowly evolving (geostrophically and hydrostatically balanced) modes of the flow.

The quality of the assimilated solutions was assessed for 4- and 8-day experiments using comparison with observations outside the respective assimilation windows. Evolution of the quantities f^n for the background and optimized solutions is shown in Fig. 6 for the 4-day assimilation experiment.

The general behavior of f is consistent with the one obtained in the 14-day experiment, showing persistently better 4dVar forecasts in temperature and the advantage of a4dVar in the salinity and velocity forecasts. The upper left panel in Fig. 6 summarizes the forecast skill and indicates that 4dVar slightly outperforms a4dVar, mostly because of the better temperature forecasts. On the other hand, the 4dVar-optimized

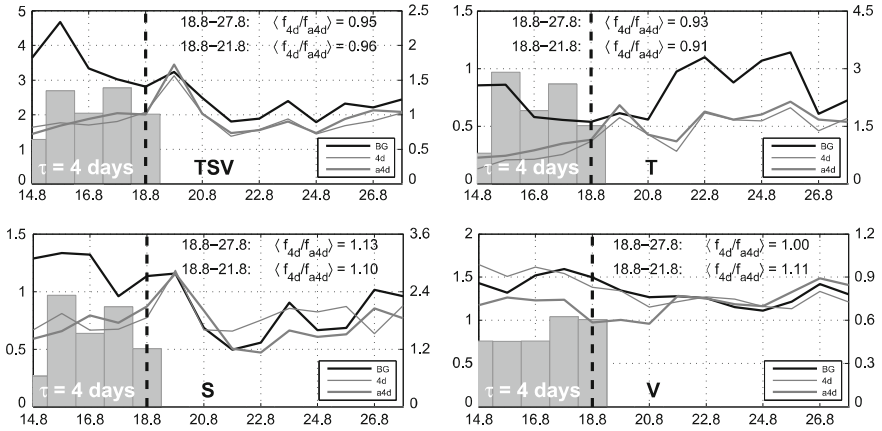

Fig. 6 Forecast skills f_s, f_v and f_t of the 4dVar and a4dVar-optimized solutions for the 4-day assimilation window. The relative number of the respective data points for each day is shown by *gray shaded rectangles*. *Vertical dashed line* show the time interval of data assimilation. Ratios of the mean f values averaged over the 3- and 9-days intervals are shown

salinity is characterized by very low forecast skill (lower left panel in Fig. 6), especially during August 21–25, when it was even farther away from the observations than the background forecast.

The 4dVar-optimized velocities show only small improvements compared to the background solution (lower right panel in Fig. 6). In contrast, the a4dVar-optimized velocities demonstrate 10–30% reduction of the model-data misfit within the assimilation window, which persists for up to three days (August 18–21) of the free model run. After August 21, the velocity mismatch of the background, a4dVar and 4dVar-optimized solutions are nearly identical. Qualitatively similar behavior of the forecast skill and its distribution among the state vector components was observed in the results of the 8-day assimilation experiment.

In general, the overall forecast skill provided by the a4dVar method appears to be comparable with that of the 4dVar (upper panel in Fig. 6), and in some aspects (such as short-term velocity forecast), the a4dVar technique provides noticeably better results. It should be noted that available observations could effectively constrain only a small part $M_d/M = 23,506/1,493,570 \sim 1.5\,\%$ of the model's degrees of freedom, so one should expect substantial differences in the small-scale structure of the optimal solutions obtained by two different methods.

3.3.2 Comparison of the Optimal Solutions

Temperature and velocity increments for the optimal states of the 14-day assimilation experiment are shown in Fig. 7. A certain coherence between the larger scale corrections to the background temperature field are clearly seen in the northern part of the model domain that is well covered by observations (cf. Fig. 1). The time-mean correlation coefficients ρ between the low-pass filtered temperature and salinity increments of the 4dVar and a4dVar solutions are 0.61 and 0.45, respectively if averaging is performed in the upper 200 m over the northern part of the domain. In the data-free region south of the 340 km mark, the correlations are substantially lower (respectively, 0.26 and 0.32) and lie below the 95 % confidence level of nonzero correlation (0.36). Similar values of ρ (0.59 and 0.32 in the northern and southern subregions, respectively) were obtained for the sea surface height field.

Velocity increments appear to have the lowest correlations among the model fields with time-averaged values of $\rho_v = 0.36, 0.27$ for the northern and southern subregions respectively. The lowest correlations ($\rho_v = 0.09, \quad \rho_T = 0.21$, and $\rho_S = 0.12$) were observed in the data-free southern subregion during the first 4 days (8/14-8/18) of the assimilation. Such incoherence between the increments is caused by excessive ageostrophic activity (lower panel in Fig. 7) of the 4dVar solution at the beginning of the assimilation window. The ageostrophic mode disappears at the later times and does not affect the cost function because the southern subregion is virtually data-free, whereas smoothness constraints are imposed on the model fields only at the initial time.

The problem could be apparently solved by introducing balance constraints (e.g., Weaver et al. 2005) into the BEC definition at $n = 0$, which may not be necessary if

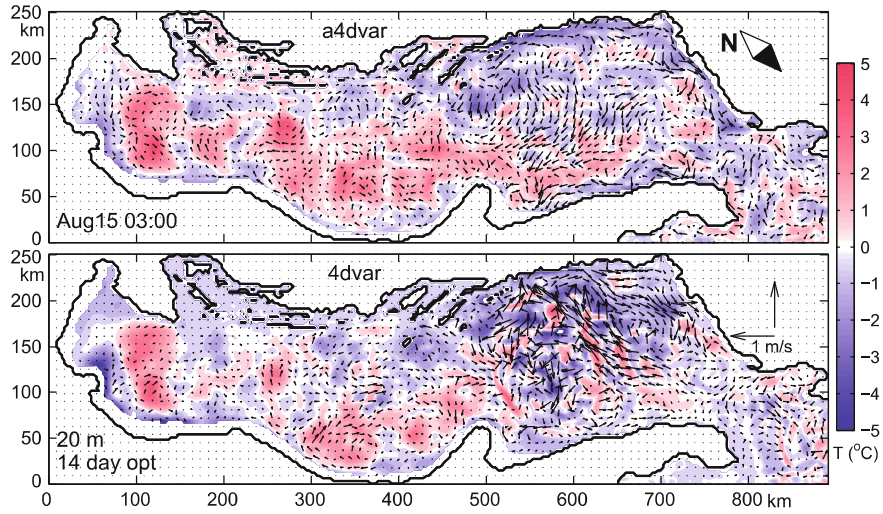

Fig. 7 Temperature and velocity differences between the background and optimized NCOM states at 20 m on August 15 03 UTC. Results of 4dVar and a4dVar optimizations are shown on the *left* and *right panels* respectively

the NCOM 4dVar were run in the weakly constrained mode, i.e., if model errors were prescribed throughout the entire assimilation window. For the purpose of comparison with a4dVar we ran the 4dVar system in the strongly constrained mode and the effect became visible after several 4dVar outer loops. It is remarkable that the a4dVar algorithm appears to be much less susceptible to excitation of the ageostrophic modes (upper panel in Fig. 7), possibly because the EOF-derived descent directions span subspaces characterized by slower time variation of the model trajectory and, therefore, tend to be closer to geostrophic and hydrostatic balance. It is quite likely that introduction of the balance constraints into $\mathbf{B}$ will certainly improve the performance of both algorithms with a potentially larger benefit for the 4dVar case.

4 a4dVar Analysis of Simulated Wave Data in the Chukchi Sea

Spectral models simulating surface gravity waves in the ocean are challenging for application of 4dVar due to complexity of their numerics and non-local nature of the observational operators. Since only a few wave models have been supplied with (incompletely) linearized codes and their adjoints, operational forecasts are still performed using sequential techniques, mostly based on optimal interpolation (OI) of the significant wave height (SWH) data from satellites. In this section, we test the

performance of a4dVar technique under the dynamical constraints imposed by a spectral wave model (WAM 1988; Monbaliu et al. 2000) and compare the results of assimilation with a sequential method.

4.1 The WAM Model and Simulated Data

The WAM model performs time integration of the balance equation describing spectral density $F(\boldsymbol{x},\boldsymbol{k},t)$ for the wave component with the wavenumber $\boldsymbol{k}=(k_x,k_y)$ at the location $\boldsymbol{x}=(x,y)$:

$$\frac{\partial F}{\partial t}+\nabla\cdot(\boldsymbol{v}F)=\mathscr{S}(F,\boldsymbol{x},\boldsymbol{k},t), \tag{15}$$

where $\mathscr{S}$ is the sum of source functions, composed primarily of wind-forced generation, dissipation and redistribution of the wave spectrum by non-linear wave-wave interactions (WAM 1988), $\nabla=\{\nabla_x,\nabla_k\}$ stands for the gradient in the horizontal and wavenumber coordinates and $\boldsymbol{v}$ is the 4-component vector of the respective wave-propagation velocities depending on the ambient current and constrained by the dispersion relationship for linear surface waves $\sigma^2=g|\boldsymbol{k}|\tanh|\boldsymbol{k}|h$, where σ is the wave angular frequency and $h(\boldsymbol{x})$ is the water depth. Given the appropriate initial/boundary conditions, ambient current and wind forcing, Eq. (15) is integrated numerically to produce evolution of the wave spectrum.

The model was configured in the domain shown in Fig. 8 with the spatial resolution of $\delta x=9$ km. There were $m_x=4412$ active grid points in horizontal and $m_k=600$ grid points (24 directions at 15° resolution and 25 logarithmically spaced frequencies between 0.0314 and 0.3091 Hz) in the wavenumber space. The total length of the state vector was $M=m_x\times m_k=2,647,200$.

Distance between the model states were assessed in terms of the correlation coefficient C and the normalized rms difference S between the spectra:

$$C(F)=\frac{\langle F_1'F_2'\rangle}{\sqrt{\langle F_1'^2\rangle\langle F_2'\rangle}};\quad S(F)=\left[\frac{\langle(F_1-F_2)^2\rangle}{\sqrt{\langle F_1'^2\rangle\langle F_2'^2\rangle}}\right]^{1/2}\qquad F'=F-\langle F\rangle \tag{16}$$

where angular brackets denote averaging in space, time, and over the wavenumbers. Similar coefficients were calculated to assess the differences between the scalar (SWH) and vector (wind speed) fields, with averaging performed just in space and time.

The general form of the cost function used in the data assimilation experiments was identical to (3) with the M-dimensional vector $\mathbf{c}=N(t_0)-N_b(t_0)$ describing the difference between the gridded model state $F(\boldsymbol{x},\boldsymbol{k},t_0)$ and the background (first guess) state $F_b(\boldsymbol{x},\boldsymbol{k},t_0)$ at the start of model integration t_0. The first term in (3) was specified by

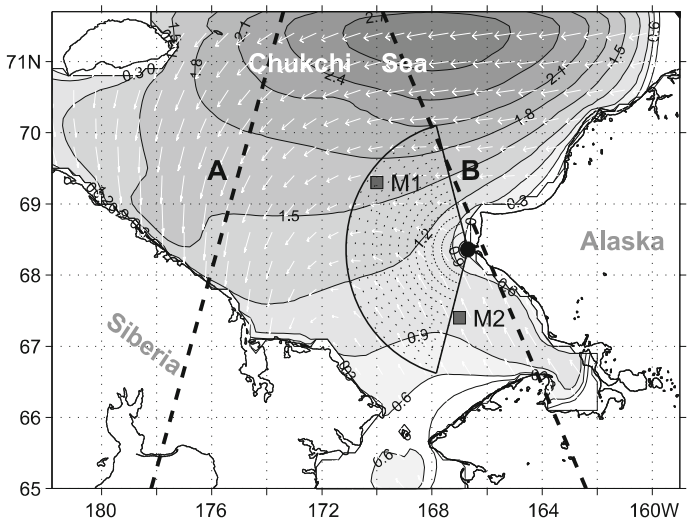

Fig. 8 Wind speed (*white arrows*) and significant wave height (contours, m) of the reference model solution at 0.00 UTC 09/20/2011 ($t = 0$). Mooring positions are shown by *black squares*. HFR location and coverage area are given by the *black circle* with a sector. SWH data are acquired along the radar beams shown by *dotted lines* within the sector. *Dashed lines* are the tested tracks of the Envisat satellite

$$\mathbf{c}^{\mathsf{T}}\mathbf{B}^{-1}\mathbf{c} = W\sum_{x}[(\mathbf{I} - a^2\Delta_x^2)\hat{Q}\mathbf{c}]^2 \tag{17}$$

where Δ_x stands for the Laplacian in horizontal coordinates and the operator $\hat{Q}$ relates squared SWH with the spectral density through the following linear relationship:

$$Q^2(\boldsymbol{x}, t) = \hat{Q}F = 16\sum_{\boldsymbol{k}} F(\boldsymbol{x}, \boldsymbol{k}, t)d\boldsymbol{k}. \tag{18}$$

Here $d\boldsymbol{k}$ denotes the grid cell area in the wavenumber space and summation is done over the entire spectral grid. In Eq. (17), the regularization weight W was chosen to be inversely proportional to the squared mean of SWH in the background solution with the proportionality coefficient $\varepsilon_x = 0.01$. By setting $a = 2\delta x$ throughout the experiments, SWH variability at spatial scales below two horizontal grid steps (18 km) was heavily penalized. In the spectral subspace, Eq. (17) defines the inverse error covariance to have only one linearly independent column (specified by the components of $\hat{Q}$). As a consequence, spectral correlations at a given point are represented by $m_k \times m_k$ correlation matrix whose elements are equal to 1 (thus implying 100 % correlation between all the spectral components). This assumption is routinely used in the sequential algorithms assimilating SWH (e.g., Wittmann and Cummings 2004).

To perform the a4dVar experiments, the reference wave field was generated by integrating WAM from the state of rest for ten days under realistic forcing by the

winds taken for the period 11–20 of September, 2011. The reference initial state $F_r(\boldsymbol{x}, \boldsymbol{k}, t)$ shown in Fig. 8 was taken at the beginning of the last 9 h of the model run (0.00 to 9.00 on 09/20/2011). Synthetic data were picked from the reference solution and then used for its reconstruction by the a4dVar method.

In this study two types of simulated data are considered: moored observations of the wave spectra and SWH measurements from coastal HF radars and satellites.

Two tested mooring sites are shown in Fig. 8. Simulated data from the moorings were generated by multiplying the reference spectrum at any moment by the random factor $1 + \varepsilon\eta$, where η is the white noise with unit variance and $\varepsilon_m = 0.01$. The observational error covariance matrices $\mathbf{R}$ for both moorings were diagonal with time-independent diagonal elements equal to $\langle \varepsilon F_r \rangle^2$. The respective observational operators $\mathbf{H}$ picked the time varying WAM spectra every 15 min from the grid point nearest to the buoy location, providing $4m_k = 2{,}400$ observations per hour.

SWH observations were simulated by integrating the true spectrum (Eq. 18) in the apexes of the grid cell containing an HFR observation point followed by linear interpolation onto that point. After that, the SWH value was contaminated by random noise with the rms variance of 30 cm. HF radar observation points were located along the beams of the radar shown in Fig. 8. The above described HFR observation operator computed SWH values along the 25 beams every 15 min, providing information to 535 model grid points within the sector shown in Fig. 8 (2,140 observations per hour).

Synthetic satellite observations of sea surface roughness provided SWH data along the Envisat tracks shown in Fig. 8 with 9 km discretization (55 and 73 points for track A and B respectively). These data were assumed instantaneous and satellite passage occurred for both tracks after 2 h of model integration. The respective observation operator was similar to the one used for HFR, except for it picked SWH values at the sequence of WAM grid points closest to the sampling points along the tracks (i.e. no spatial interpolation was used). Satellite SWH observations were contaminated similarly to HFRs with the rms error variance of 30 cm.

The background model trajectory was obtained as follows: The reference solution was averaged in time and space and the resulting spatially homogeneous spectrum was used as initial condition for the background run. The run was forced by the winds which were different from those forcing the true solution. First, the true winds were horizontally smoothed to mimic the errors typical for reanalysis winds from meteorological centers that are usually available at a coarser (0.25−1°) resolution and have to be interpolated on the fine resolution grid of a regional wave model. In the case considered, the smoothing was done by the isotropic Gaussian filter with the half-width of 25 km. After smoothing, the winds were rotated 35° counterclockwise to increase their distance from the reference vectors to $S_{wind} = 0.67$. The larger distance from the true forcing was needed for better assessment of the observation impact on the reconstruction of initial conditions, whose signature usually persists for 3–5 h in a typical wave model integration.

Synthetic observations of SWH and wave spectra were assimilated into WAM using the a4dVar technique described in Sect. 3. The WAM model was constrained by data during the first three hours of model integration and then integrated for six

hours to assess the improvement of the forecast skill. Performance of the method was quantified by calculating correlation coefficients C and normalized rms deviation S (Eq. 16) between the optimized and true solutions. These quantities were computed with time averaging over three time intervals: 0–3 h (assimilation period), and two forecast periods of 3–6, and 6–9 h.

On each a4dVar iteration five SDs were extracted from the EOF analysis of the 3 h model run constrained by the data. The ensemble model runs (p. 3 in the layout of Sect. 3.2) were executed in parallel and required 62 s of wall time per a4dVar iteration on five processors of the 2.3 GHz cluster.

4.2 Comparison with Sequential Method

To compare the a4dVar results with the traditional OI method, we used the 2d OI approach (Wittmann and Cummings 2004; Waters et al. 2013) in application to the SWH data: at the observation times the WAM model state was sequentially updated by the OI analysis of the SWH field, which was projected on the spectral components by multiplying the spectrum at a grid point by the ratio of the updated to predicted SWH values. The OI algorithm was configured with the same background error covariance $\mathbf{B}$, $\mathbf{R}_n$, $\mathbf{H}_n$ and using the same reference and background solutions as the a4dVar method.

A series of OI and a4dVar experiments were conducted, involving assimilation of the data from five sources and their combinations: high-frequency radar (hereinafter denoted by HF), two moorings (a4dVar analyses only, locations shown in Fig. 8) and two Envisat tracks (A and B, Fig. 8). For comparison purposes, we conducted similar experiments with OI method assimilating only SWH data from satellites and/or HF radar. In the description below, these experiments are abbreviated by oHFA(B) and oHF respectively. With the exception of satellite tracks, all a4dVar assimilation experiments demonstrated significant improvement of the model state in terms of its proximity to the reference solution. The stopping criterion for optimization was reduction of the cost function gradient 1000 times, which usually occurred after 80–100 iterations. By that time the cost function was typically reduced 2–3 times.

Maps of deviations from the reference solution of the spatially averaged background and optimized spectra at $t = 0$ are shown in Fig. 9. In most of the a4dVar experiments, the initial error has been reduced to the values compatible with the wind forcing errors. The only exceptions were the results of optimal interpolation (Fig. 9b) and of the a4dVar assimilation of SWH data from a single satellite track (not shown): in these cases the optimized spectrum was only slightly different from the one produced by the background solution. For the OI case such a small correction can be explained by the fact that SWH data are weakly constrained by dynamics and can barely affect the shape and location of the spectra because they provide information only on their mean magnitude at a geographical position. Small spectral improvement of the a4dVar experiments with a single satellite track could be attributed to the small amount of data (55 SWH observations). As a consequence, the cost func-

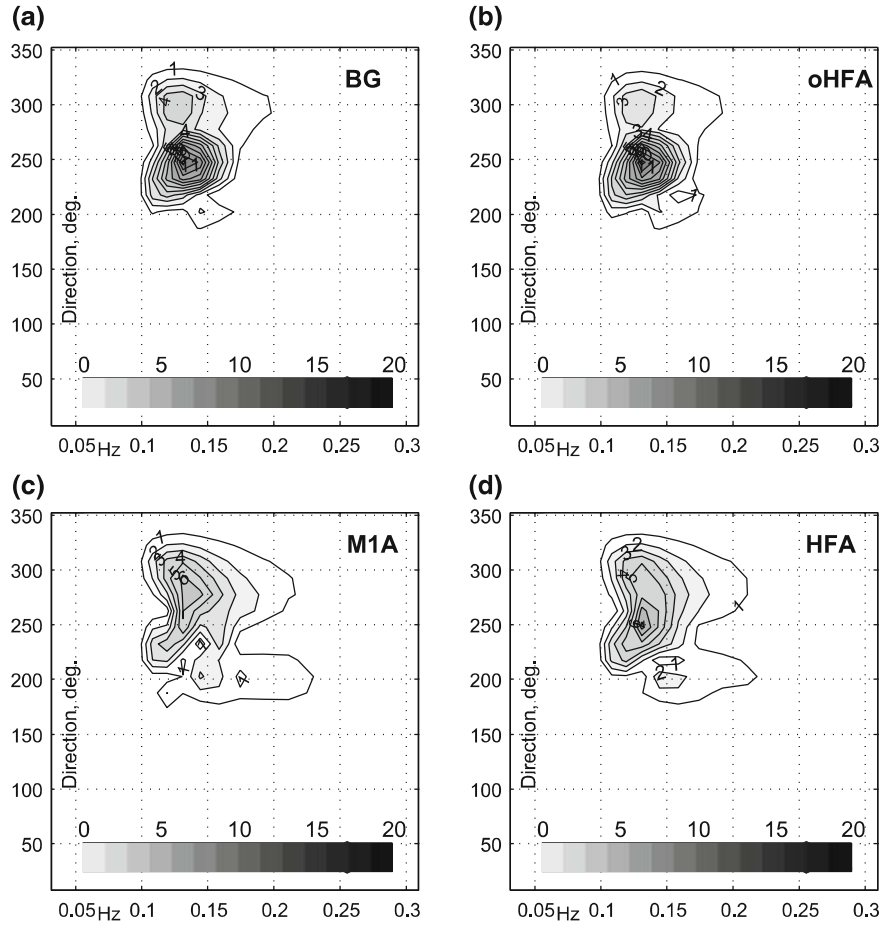

Fig. 9 Absolute difference between the horizontally averaged reference spectrum at $t = 0$ and **a** background spectrum, **b** oHFA-optimized, **c** M1A-optimized, and **d** HFA-optimized spectra

tion is dominated by the regularization term, which implies 100 % correlations in spectral space and is therefore capable of adjusting only the spectral magnitude.

These properties of the above mentioned assimilated solutions translate into their lower spectral forecast skill shown in Table 1, which also includes spectral errors from the other assimilation experiments. Abbreviations in the header of the Table correspond to the types of data used in the experiment (e.g., HFA corresponds to assimilation of the HF data and SWH data from the Envisat track A).

Direct measurement of the spectra by a single mooring (7,200 observation points, columns M1,M2) also provide only a moderate increase of the correlation coefficients C_{03} to 0.52 and decrease of S_{03} to 0.86 as compared to the background (BG) solution. This can be partly explained by the fact that assimilated spectra occupy a

Table 1 Normalized rms distances S and correlations C between the optimized and true solutions for the experiments with various types of data. Subscripts 03, 36 and 69 correspond to time averaging between 0–3, 3–6, and 6–9 h of model integration

	BG	HF	HFA	HFB	oHFA	M1	M1A	M1B	M2	M2A	M2B	M12	A	B
C_{03}	0.47	0.77	0.75	0.76	0.48	0.53	0.70	0.71	0.52	0.65	0.69	0.71	0.47	0.48
S_{03}	0.89	0.65	0.66	0.64	0.87	0.85	0.71	0.70	0.87	0.76	0.75	0.73	0.89	0.88
C_{36}	0.48	0.72	0.72	0.76	0.49	0.47	0.70	0.70	0.50	0.69	0.69	0.67	0.48	0.48
S_{36}	0.87	0.69	0.68	0.65	0.86	0.87	0.71	0.71	0.87	0.72	0.72	0.75	0.88	0.87
C_{69}	0.59	0.71	0.70	0.75	0.59	0.50	0.68	0.69	0.57	0.67	0.68	0.68	0.57	0.58
S_{69}	0.80	0.70	0.70	0.67	0.79	0.86	0.73	0.72	0.86	0.73	0.73	0.73	0.79	0.78

small part of spectral domain (at most 15–20 %, Fig. 9). As a consequence, the effective number of observations with useful (non-zero) information on the state of the wave field should be reduced 5–7 times down to ∼1,500 data points on the total, which is compatible, by the way, to assimilating 3–7 spectral moments. Besides, mooring data do not provide any information on the spatial variability of the spectra, which appears to be crucial for the successful recovery of the reference state.

In that respect, it is remarkable that adding much less numerous satellite data to moored spectral observations improves the performance of the assimilation system considerably. Combining moored and satellite data provides 30–40 % growth of the correlation coefficients and 20–25 % drop of the normalized standard deviations from the reference spectrum (compare columns M1 and M2 with columns M1A(B) and M2A(B)). At the same time, Satellite SWH data do not add much new information to that containing in HFR observations (cf. columns HF and HFA(B)), which monitor the same integral quantity for the whole assimilation period (3 h) and cover a significant part of the model domain (Fig. 8).

Importance of the spatial coverage by observations is confirmed by the result of the experiment with assimilation of the spectra from two moorings: The values of C and S in this case demonstrate a considerable improvement and become compatible (column M12 in Table 1) with those achieved with the joint assimilation of spectra from the single mooring and satellite SWH data (columns M1A(B) and M2A(B)).

Inspection of Table 1 also shows that information from track A increases the efficiency of assimilating spectra from moorings, but to somewhat lesser extent than track B. This phenomenon can be partly explained by the fact that track A does not cover the region of the highest SWH and, therefore, provides less information on the magnitude of spatial variability of the wave field. Similarly, assimilation of the M2 data appears to be slightly less efficient than M1, which can be partly attributed to M2 position at the periphery of the domain.

Table 1 also indicates that instantaneous Envisat observations on a regional scale cannot provide a significant improvement to the background state, if they are not accompanied by continuous in situ measurements. At the same time, satellite data

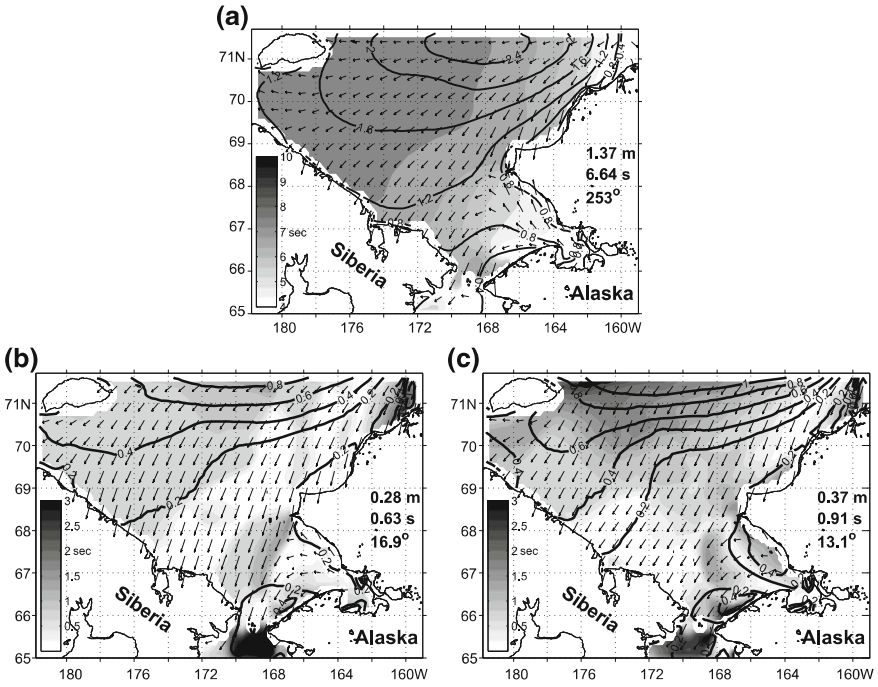

Fig. 10 Time-averaged (3–9 h) values of SWH (contours, m), peak period (*shading*, seconds) and wave direction (*arrows*) of the true state **a** and respective absolute differences of the HF-optimized **b** and oHF **c** solutions from the true state. The domain-averaged values of the fields are shown on the *right*

become quite valuable in complementing observations if the wave conditions are measured by a single mooring.

The forecast errors provided by the HFA data assimilation using OI and a4dVar techniques and averaged over the period 3–9 h are compared in Fig. 10 in terms of the horizontal distributions of the SWH, peak period and wave direction errors. It is seen, that a4dVar technique provides 30–50 % better forecast skill in terms of the SWH (0.28 vs 0.37 m) and peak period (0.63 vs 0.91 s). Although discrepancies in the peak period near the southern and eastern boundaries are comparable in both solutions, the a4dVar method demonstrates a significant advantage over OI in the northern Chukchi Sea and south of Cape Hope resulting in approximately 10 cm smaller SWH errors throughout the entire domain. A local maximum in the a4dVar peak period errors is also observed southwest of Cape Hope (Fig. 10b), that can be partly explained by a sharper gradient in the peak period field of the true solution (Fig. 10a).

The OI solution demonstrates a slightly better skill in forecasting the wave direction (the mean difference of 13.1° vs 16.9°). However, in the OI assimilation experiments with other types of SWH data this number varied within 13−13.3° and was quite close to the respective characteristic (13.2°) of the background solution.

In general, our experiments have shown that the OI method tends to improve the amplitude of the spectrum, and has only a slight impact on its shape and position in the frequency-direction coordinates. In contrast, a4dVar technique is capable of improving these characteristics as well, since it performs optimization along the most persistent dynamical modes of the governing equation (15). This important property of the a4dVar algorithm provides a significantly better approximation of the reference solution and improved forecast skill.

5 Summary and Discussion

In this chapter we have shown feasibility of the a4dVar technique (Yaremchuk et al. 2009) in realistic applications and compared its performance with the observation space 4dVar and OI methods. It was shown that the a4dVar approach is capable of producing optimized solutions of similar quality to 4dVar with comparable computational expense. It was also found that the a4dVar technique is less susceptible to excitation of ageostrophic modes in the data-free regions if balance constraints are not imposed on the background error covariances.

The a4dVar technique employs square root factorization the inverse BEC and the possibility of inexpensive evaluation of the product $\tilde{\mathbf{H}}^{1/2}\delta\mathbf{c}$ during the integration of the ensemble of perturbed model trajectories. The technique of Hessian factorization was first proposed by Županski (2005) in the framework of minimizing the cost function within the subspace spanned by the ensemble members. It was later extended in Yaremchuk et al. (2009) to heuristic BEC models coupled with iterative ensemble updates produced by projections of the model-data misfits on a suitable "smooth" manifold generated by the low-pass filtering operators $\mathbf{M}^n$ or $\mathbf{B}$.

Our experience shows that there exists a considerable freedom in generating the SDs as long as they are kept being spatially smooth and $\tilde{\mathbf{H}}$-orthogonal. In particular, selecting the ensembles as eigenvectors of $\mathbf{B}$ in the decreasing order of their eigenvalues proves to be equally efficient, at least in the simple linear setting considered in Sect. 2. In that respect, there is a considerable similarity between the a4dVar and the adjoint-free 4dEnVar method (Liu et al. 2008), which explicitly looks for optimized solution in the range of the localized approximation to the background error covariance. However, the 4dEnVar uses the ensemble to approximate the cost function gradient which is then used in the iterative optimization, whereas a4dVar directly employs the ensemble members to minimize the cost function in the respective subspace.

The ultimate goal of a search algorithm is to rapidly gain information on the Hessian structure, which helps to find the SD $\mathbf{s} = \tilde{\mathbf{H}}^{-1}\mathbf{b}$ towards the cost function minimum (note that $\mathbf{s}$ is often nearly orthogonal to the local gradient). In that respect, ensemble methods can offer a significant advantage in their ability to perform *parallel* searches in *multiple SDs*, which can be competitive with the adjoint-based methods even in some cases where individual SDs may not appear to be as efficient as steepest descent or conjugate gradient directions.

Regarding linearization issues, a state-of-the-art GCM code is never fully differentiable and its (always approximate) adjoint usually requires several times more CPU/memory resources than the direct model run. This observation indicates that a4dVar approach could be even competitive with 4dVar even in terms of the total CPU time at small ensemble sizes. The present chapter demonstrates this compatibility in both a real-life scenario and a simplified linear application.

The a4dVar technique can be developed further by introducing flow-dependent covariances and better restricting the SDs to the slow-evolving (geostrophically and hydrostatically balanced) manifold. In application to atmospheric and oceanic modeling, the BEC matrix is can easily incorporate these balance constraints by representing the state vector in the form

$$\mathbf{x} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{L} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \tag{19}$$

where $\mathbf{L}$ is the balance operator (Weaver et al. 2005), $\mathbf{x}_{1,2}$ are the unbalanced components of the state vector, and $\mathbf{I}_{1,2}$ are the identity matrices of the respective sizes. Under these constraints, $\mathbf{B}^{-1} = \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle^{-1}$ in the a4dVar formulation will take the form

$$\mathbf{B}_{bal}^{-1} = \begin{bmatrix} \mathbf{B}_1^{-1} + \mathbf{L}^{\mathrm{T}}\mathbf{B}_2^{-1}\mathbf{L} & -\mathbf{L}^{\mathrm{T}}\mathbf{B}_2^{-1} \\ -\mathbf{B}_2^{-1}\mathbf{L} & \mathbf{B}_2^{-1} \end{bmatrix}, \tag{20}$$

where $\mathbf{B}_1^{-1}$ and $\mathbf{B}_2^{-1}$ are the inverse covariances of $\mathbf{x}_1$ and $\mathbf{x}_2$. Further improvements can be made by replacing the Laplacian in Eq. (7) with a more general expression (e.g., Weaver and Mirouze 2012; Yaremchuk and Nechaev 2013) introducing flow dependent structure into $\mathbf{B}_{1,2}^{-1}$ while keeping them square root factorisable. Note that spectral analysis of the background error covariance could be efficiently performed prior to the assimilation.

Alternatively, flow-dependence and cross-correlations could be introduced into BEC through its representation by the localized external ensembles, as it is done in 4dEnVar. This will require $\mathbf{B}$-preconditioning of the control variables, which will bring the method closer to the observation space 4dVar. In that respect, it is interesting to note a certain similarity between the a4dVar and observation space 4dVar: The latter method explicitly computes the Hessian projection on the observation space (representer matrix), whose computation is efficiently parallelized between M_d processors, making the method competitive with a4dVar in terms of scalability on massively parallel computers. This property brings observation space 4dVar

closer to the family of optimization algorithms capable of taking the advantage of massive parallelism. In our vision, such algorithms are getting higher priority under the current "parallelization trend" in the development of computer technologies. In a sense, the present situation is somewhat similar to the situation 30 years ago when the adjoint methods started coming into practice in response to the rapid increase of computer speed and memory.

In terms of the computational expense, the tested a4dVar technique appears comparable to 4dVar, mostly because of the excessive computational cost of tangent linear and adjoint codes that were, on average, several times more expensive than a direct run of the parent nonlinear model (which is a typical situation with state-of-the-art OGCMs, e.g., Oldenborgh et al. 1999). On massively parallel machines, the advantage of a4dVar will be more noticeable due to the limited parallel scalability of an OGCM code, be it original, adjoint, or tangent linear.

An important issue with the a4dVar technique is its extension to optimization of other sets of variables that may control the model trajectory, such as surface forcing fields. One of the possible solutions in this case augments the search subspaces (ocean model states) by the leading EOFs of the surface forcing error fields. This will require a better knowledge of error statistics of the atmospheric model used to force the ocean in a particular application. In view of recent rapid development of the observational systems and data acquisition techniques in the atmosphere, the issue of accessibility to the above mentioned statistics seems likely to be resolvable in the near term. Moreover, the a4dVar technique appears to be even more suitable for coupled ocean-atmosphere systems, where external forcing errors tend to play a lesser role at the time scale of a typical assimilation window.

A much larger computational advantage is evident when considering the wall time in a massively parallel environment, which formally allows a4dVar to search over multiple directions at a fraction of the wall time used by 4dVar to generate a steepest descent direction. In fact, in the experiments reported in Sect. 3, one a4dVar run was executed almost five times faster if all the ensemble members were run on separate nodes. This property of the a4dVar approach gives good prospects for its further development in sync with other types of ensemble data assimilation techniques that are based on relaxed communication requirements between processors. In our vision, rapidly decreasing prices of the massively parallel computers make finite differentiation in functional spaces more affordable, favoring development of ensemble methods of data assimilation, while investment in the development and maintenance of linearized codes and their adjoints may gradually become less practical.

Acknowledgements This study was supported by the Office of Naval Research (Program element 0602435N) as part of the project "Adjoint-free 4dVar for Navy ocean models". Partial support from NSF grants PLR-1107925, PLR-1203740, DMS-1217156 and Einstein Foundation of Berlin is also acknowledged.

Appendix

The a4dVar method utilizes the technique employed by Županski (2005) in the Maximum Likelihood Ensemble Filter, which is based on the explicit inversion of the Hessian matrix in the subspace spanned by the model perturbations. In view of the definition (7), $\mathbf{B}^{-1/2}$ can be explicitly represented using the expression for the square root of the inverse error covariance:

$$\mathbf{B}^{-1/2} = \mathbf{V}^{-1}(\mathbf{I} - \frac{b^2}{2}\Delta) \tag{21}$$

which allows a symmetric Hessian factorization

$$\tilde{\mathbf{H}} = \tilde{\mathbf{H}}^{\mathrm{T}/2}\tilde{\mathbf{H}}^{1/2}, \tag{22}$$

where

$$\tilde{\mathbf{H}}^{\mathrm{T}/2} = \begin{bmatrix} \mathbf{B}^{-1/2} & \mathbf{H}_0 & \mathbf{H}_1\mathbf{M}^1 & \dots & \mathbf{H}_N\mathbf{M}^N \end{bmatrix} \tag{23}$$

is the Hessian square root.

For sufficiently small perturbations $\delta\mathbf{c}_m = \varepsilon\mathbf{p}_m$, perturbations of the the auxiliary vector

$$\delta\mathbf{Y}_m = \tilde{\mathbf{H}}^{1/2}\delta\mathbf{c}_m \tag{24}$$

are linear in $\delta\mathbf{c}_m$, so that computation of the dot products between the vectors $\delta\mathbf{Y}_m$ provides the inner product in the control space associated with the Hessian matrix

$$\delta\mathbf{Y}_1^{\mathrm{T}}\delta\mathbf{Y}_2 = \delta\mathbf{c}_1^{\mathrm{T}}\tilde{\mathbf{H}}\delta\mathbf{c}_2 = \langle\delta\mathbf{c}_1, \delta\mathbf{c}_2\rangle_{\tilde{\mathbf{H}}}, \tag{25}$$

which can be used for $\tilde{\mathbf{H}}$-orthogonalization of the search subspaces of the a4dVar algorithm.

We seek the optimal correction of the control variable $\mathbf{c}$ in the search subspace $\mathbb{S}$ spanned by $\mathbf{p}_m$:

$$\mathbf{c} \leftarrow \mathbf{c} + \sum_{l=1}^{m_s} s_l\mathbf{p}_l,$$

where the coefficients s_l satisfy for $m = 1, 2, \dots, m_s$,

$$\mathbf{p}_m^{\mathrm{T}}\left(\tilde{\mathbf{H}}\Big(\mathbf{c} + \sum_{l=1}^{m_s} s_l\mathbf{p}_l\Big) - \mathbf{b}\right) = 0. \tag{26}$$

This constitutes a Ritz-Galerkin projection of the normal system (4) to the search subspace, $\mathbb{S}$. Rearranging, we obtain the linear system of m_s equations in the m_s unknowns $s_1, s_2, \ldots, s_{m_s}$:

$$\sum_{l=1}^{m_s} \mathbf{p}_m^\top \tilde{\mathbf{H}}\, \mathbf{p}_l s_l = \mathbf{p}_m^\top (\mathbf{b} - \tilde{\mathbf{H}}\mathbf{c}). \tag{27}$$

Substituting $\mathbf{p}_m = \delta\mathbf{c}_m/\varepsilon$ into (27), multiplying by ε^2, using (22) and (24) yields

$$\sum_{l=1}^{m_s} \delta\mathbf{Y}_m^\top \delta\mathbf{Y}_l s_l = \varepsilon \delta\mathbf{c}_m^\top (\mathbf{b} - \tilde{\mathbf{H}}\mathbf{c}). \tag{28}$$

The right-hand side of (A.7) cannot be computed directly because evaluation of $\mathbf{b} - \tilde{\mathbf{H}}\mathbf{c}$ requires the adjoint code (Eq. 5). Nonetheless, for each m, $\delta\mathbf{c}_m^\top(\mathbf{b} - \tilde{\mathbf{H}}\mathbf{c})$ can be calculated directly from the variations of the cost function $\delta J_m = J(\mathbf{c} + \delta\mathbf{c}_m) - J(\mathbf{c})$ induced by $\delta\mathbf{c}_m$:

$$\begin{aligned} \delta J_m &= \frac{1}{2}\delta\mathbf{c}_m^\top \tilde{\mathbf{H}} \delta\mathbf{c}_m + \delta\mathbf{c}_m^\top \left(\tilde{\mathbf{H}}\mathbf{c} - \mathbf{b}\right) \\ &= \frac{1}{2}\delta\mathbf{Y}_m^\top \delta\mathbf{Y}_m - \delta\mathbf{c}_m^\top \left(\mathbf{b} - \tilde{\mathbf{H}}\mathbf{c}\right). \end{aligned} \tag{29}$$

Thus, the coefficients for the optimal correction of the control variable $\mathbf{c}$ within the search subspace $\mathbb{S}$ are given as the solution to a linear system posed in terms of the quantities δJ_m and $\delta\mathbf{Y}_m$ computed by the a4dVar algorithm:

$$\sum_{l=1}^{m_s} \delta\mathbf{Y}_m^\top \delta\mathbf{Y}_l s_l = \varepsilon \left(\frac{1}{2}\delta\mathbf{Y}_m^\top \delta\mathbf{Y}_m - \delta J_m\right). \tag{30}$$

In the $\tilde{\mathbf{H}}$-orthonormal coordinate system $\delta\mathbf{Y}_m^\top \delta\mathbf{Y}_m = \varepsilon^2$, and Eq. (30) are simplified to

$$s_l = \sum_m \alpha_{lm} \left(\frac{\varepsilon}{2} - \frac{\delta J_m}{\varepsilon}\right), \tag{31}$$

where α_{lm} are the matrix elements of the linear transformation of the original basis $\delta\mathbf{c}_m$ that are obtained in the orthogonalization process.

For a differentiable numerical model and sufficiently small ε, the quadratic term in the right hand side of (30) is negligible. In the experiments resported in Sect. 3 we kept it in place since the value of ε was close to 0.01 and could not be reduced any further without affecting the rate of convergence. The relatively large limit on the value of ε was caused by a number of factors deteriorating the linear dependence between the magnitude of the model perturbations and ε. These factors include rounding errors (especially for temperature and salinity in the upper layers),

non-differentiable operators in the model code, particularly at the open boundary, and small-scale instabilities of the flow, especially prominent in the experiments with the 14-day assimilation window.

References

Anderson JL, Hoar T, Raeder K, Liu H, Collins N, Torn R, Arellano A (2009) The data assimilation research testbed: a community facility. Bull Am Meteorol Soc 90:12831296

Barron CN, Kara AB, Martin PJ, Rhodes RC, Smedstad LF (2006) Formulation, implementation and examination of vertical coordinate choices in the global Navy Coastal Ocean Model (NCOM). Ocean Modell 11:347375

Bennett AF (2002) Inverse modeling of the ocean and atmosphere. Cambridge University Press, pp 234. ISBN 0-521-81373-5

Burrage DM, Book JW, Martin PJ (2009) Eddies and filaments of the Western Adriatic Current: Analysis and prediction, J Mar Syst, 78, S205–S226

Clayton AM, Lorenc AC, Barker DM (2013) Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q J R Meteorol Soc. doi:10.1002/qj.2054

Desroziers G, Camino J-T, Loik Berre (2014) 4dEnVar: link with 4D state formulation of variational assimilation and different possible implementations. Q J R Meteorol Soc 140:2097–2110

Fairbairn D, Pring SR, Lorenc AC, Roulstone I (2014) A comparison of 4dVar with ensemble data assimilation methods. Q J R Meteorol Soc 140:281–294

Hamill TM, Whitaker JS, Snyder C (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon Weather Rev 129:2776–2790

Hoteit I (2008) A reduced-order simulated annealing approach for four-dimensional variational data assimilation in meteorology and oceanography. Int J Numer Methods Fluids 58:1181–1199

Hoteit I, Luo X, Pham DT (2012) Particle Kalman filtering: a nonlinear Bayesian framework for ensemble Kalman filters. Mon Weather Rev 140:528–542

Hoteit I, Hoar T, Gopalakrishnan G, Anderson J, Collins N, Cornuelle B, Kohl A, Heimbach P (2013) A MITgcm/DART ocean prediction and analysis system with application to the Gulf of Mexico. Dyn Atmos Oceans 63:1–23

Kuhl DD, Rosmond TE, Bishop CH, McLay J, Baker N (2013) Comparison of hybrid ensemble/4dVar and 4dVar within the NAVDAS-AR data assimilation framework. Mon Weather Rev 141:2740–2758

Liu C, Xiao Q, Wang B (2009) An ensemble-based four-dimensional variational data assimilation scheme. Part II: observing system simulation experiments with advanced research WRF (ARW). Mon Weather Rev 137:1687–1704

Liu C, Xiao Q, Wang B (2013) An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with advanced research WRF using real data. Mon Weather Rev 141:2721–2739

Liu C, Xiao Q, Wang B (2008) An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. Mon Weather Rev, 136, 3363–3373

Martin PJ, Book JW, Burrage DM, Rowley CD, Tudor M, (2009) Comparison of model-simulated and observed currents in the central Adriatic during DART, J Geophys Res, 114, C01S05. doi:10.1029/2008JC004842

Martin PJ (2000) A description of the Navy Coastal Ocean model version 1.0, NRL Rep. NRL/FR/7322 00-9962, Nav Res Lab, Stennis Space Cent, MS pp 42

Menemenlis D, Wunsch C (1997) Linearization of an oceanic general circulation model for data assimilation and climate studies. J Atmos Ocean Technol 14:1420–1443

Mirouze I, Weaver A (2010) Representation of correlation functions in variational data assimilation using an implicit diffusion operator. Q J R Meteorol Soc 136:14211443

Monbaliu J, Padilla-Hernndez R, Hargreaves JC, Carretero Albiach JC, Luo W, Sclavo M, Gnther H (2000) The spectral wave model WAM, adapted for applications with high spatial resolution. Coast Eng 41:41–62

Moore AM, Arango HG, Broquet G, Powell BS, Weaver AT, Zavala-Garay J (2011) The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I system overview and formulation. Prog Ocean 91(1):34–49

Ngodock H, Carrier M (2014) A 4dVar system for the Navy Coastal Ocean Model. Part I: system description and assimilation of synthetic observations in Monterey Bay. Mon Weather Rev 142(6):2085–2107

Oldenborgh GJ, Burgers G, Venzke S, Eckart C, Giering R (1999) Tracking down the ENSO delayed oscillator with an adjoint OGCM. Mon Weather Rev 127:1477–1495

Panteleev G, Yaremchuk M, Rogers E (2015) Adjoint-free variational data assimilation into a regional wave model. J Atmos Ocean Technol, 32. (in press)

Qui C, Shao A, Wei L (2007) Fitting model fields to observations by using singular value decomposition: an ensemble-based 4dVar approach. J Geophys Res 112:D11105. doi:10.1029/2006JD007994

Rosmond T, Xu L (2006) Development of the NAVDAS-AR: non-linear formulation and outer loop tests. Tellus 58A:45–58

Stammer D, Wunsch C (1996) The determination of the large-scale circulation of the Pacific Ocean from satellite altimetry using model Green's functions. J Geophys Res 101:18409–18432

The WAMDI Group (1988) The WAM model—a third generation wave prediction model. J Phys Ocean 18:1775–1810

Tian X, Xie Z (2012) Implementations of a square-root ensemble analysis and a hybrid localization into the POD-based ensemble 4dVar. Tellus A 64:1–10. doi:10.3402/tellusa.v64i0.18375

Trevisan A, DIsidoro M, Talagrand O (2010) Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. Q J R Meteorol Soc 136:487–496

Waters J, Wyatt LR, Wolf J, Hines A (2013) Data assimilation of partitioned HF radar wave data into Wavewatch III. Ocean Modell 72:17–31

Weaver AT, Deltel C, Machu E, Ricci S, Daget N (2005) A multi-variate balance operator for variational data assimilation. Q J R Meteorol Soc 131:3605–3625

Weaver AT, Mirouze I (2012) On the diffusion equation and its application to isotropic and anisotropic correlation modeling in variational assimilation. Q J R Meteorol Soc 138. doi:10.1002/qj.1953

Weaver AT, Courtier P (2001) Correlation modelling on a sphere using a generalized diffusion equation. Q J R Meteorol Soc 127:18151846

Wittmann PA, Cummings JA (2004) Assimilation of altimeter wave measurements into WAVEWATCH III. In: 8th International Workshop on Wave Hindcasting and Forecasting, North Shore, Oahu, Hawaii, 14–19 November 2004

Xu L, Rosmond T (2004) Formulation of the NRL atmospheric variational data assimilation system—accelerated representer (NAVDAS-AR), NRL/MR/7532-36. Naval Research Laboratory, pp 28

Xu L, Rosmond T, Daley R (2005) Development of the NAVDAS-AR: formulation and initial tests of the linear problem. Tellus 57A:546–559

Yaremchuk M, Nechaev D, Panteleev G (2009) A method of successive corrections of the control subspace in the reduced-order variational data assimilation. Mon Weather Rev 137:2966–2978

Yaremchuk M, Carrier M, Smith S, Jacobs G (2013) Background error correlation modeling with diffusion operators. In: Park SK, Xu L (eds) Data Assimilation for Atmospheric, Oceanic and Hydrological Applications, vol. 2, Springer, pp 177–203

Yaremchuk M, Nechaev D (2013) Covariance localization with the diffusion-based correlation models. Mon Weather Rev 141:848–860

Yaremchuk M, Sentchev A (2012) Multi-scale correlation functions associated with polynomials of the diffusion operator. Q J R Meteorol Soc 138:1948–1953

Yaremchuk M, Smith S (2011) On the correlation functions associated with polynomials of the diffusion operator. Q J R Meteorol Soc 137:1927–1932

Zhang F, Zhang M, Hansen JA (2009) Coupling ensemble Kalman filter with four-dimensional variational data assimilation. Adv Atmos Sci 26:19

Zhang M, Zhang F (2012) E 4dVar: coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. Mon Weather Rev 140:587–600

Županski M (2005) Maximum likelihood ensemble filter: theoretical aspects. Mon Weather Rev 133:1710–1726